Enhancing Heatwave Predictability in Bangladesh through Deep Learning Approaches

A thesis
Submitted in partial fulfillment of the requirements for the Degree of
Bachelor of Science in Computer Science and Engineering

Submitted by

Sumaiya Siddiqua Mumu	**190204040**
Syeda Samia Sultana	**190204048**
Noshin Nawar Neha	**190204025**
Rafid Reezwan Fahim	**190204082**

Supervised by

Dr. Md. Shahriar Mahbub

Department of Computer Science and Engineering
Ahsanullah University of Science and Technology

Dhaka, Bangladesh

April 03, 2024

CANDIDATES’ DECLARATION

We, hereby, declare that the thesis presented in this report is the outcome of the investigation performed by us under the supervision of Dr. Md. Shahriar Mahbub, Department of Computer Science and Engineering, Ahsanullah University of Science and Technology, Dhaka, Bangladesh. The work was spread over two final year courses, CSE4100: Project and Thesis I and CSE4250: Project and Thesis II, in accordance with the course curriculum of the Department for the Bachelor of Science in Computer Science and Engineering program.

It is also declared that neither this thesis nor any part thereof has been submitted anywhere else for the award of any degree, diploma or other qualifications.

Sumaiya Siddiqua Mumu
190204040

Syeda Samia Sultana
190204048

Noshin Nawar Neha
190204025

Rafid Reezwan Fahim
190204082

CERTIFICATION

This thesis titled, **“Enhancing Heatwave Predictability in Bangladesh through Deep Learning Approaches”**, submitted by the group as mentioned below has been accepted as satisfactory in partial fulfillment of the requirements for the degree B.Sc. in Computer Science and Engineering in April 03, 2024.

Group Members:

Sumaiya Siddiqua Mumu	**190204040**
Syeda Samia Sultana	**190204048**
Noshin Nawar Neha	**190204025**
Rafid Reezwan Fahim	**190204082**

Dr. Md. Shahriar Mahbub
Professor & Supervisor
Department of Computer Science and Engineering
Ahsanullah University of Science and Technology

Dr. Md. Shahriar Mahbub
Professor & Head
Department of Computer Science and Engineering
Ahsanullah University of Science and Technology

ACKNOWLEDGEMENT

First and foremost, We are grateful to the Almighty Allah for the good health and well being that were necessary to complete this thesis work. Then we have to thank our thesis supervisor, Dr. Md. Shahriar Mahbub - without whose encouragement, this thesis would have never been accomplished. His constant support, understanding and constructive critiques over the final year have enriched our thesis work to a great extent.

We take this opportunity to express gratitude to all the respected faculty members of our department for their help and support. Also, we thank our parents for the constant encouragement, support and attention. We are grateful to everyone who helped us in every possible ways to make our thesis fruitful.

Lastly, we want to show our earnest gratitude towards S.M. Quamrul Hassan(Meteorologist, Storm Warning Center) and Dr. Muhammad Abul Kalam Mallik (Meteorologist, Storm Warning Center) from Bangladesh Meteorological Department for their valuable support.

Dhaka
April 03, 2024

Sumaiya Siddiqua Mumu

Syeda Samia Sultana

Noshin Nawar Neha

Rafid Reezwan Fahim

ABSTRACT

In recent years, Bangladesh has experienced some of the highest temperatures in six decades as an outcome of the enormous heat wave that is scorching across Asia. In sensitive locations such as Bangladesh, heatwaves have a substantial influence on public health, agricultural production, energy needs, and overall societal resilience [1]. Therefore, a thorough understanding and precise forecast of heatwave occurrences are essential for proactive measures. Predicting upcoming heatwave days may help a lot of people and different sectors in developing countries. For our research, the dataset has been collected from the Bangladesh Meteorological Department(BMD). The dataset contains 25 years of historical day-to-day weather data of different districts of Bangladesh. The districts are Dhaka, Rajshahi, Bogra, Dinajpur, Khulna and Jessore which are some of the most heatwave-prone districts of Bangladesh. Weather data for March, April and May were selected as they are the most heatwave-prone months in Bangladesh. This study has looked into different deep learning-based models to understand their ability to predict heat waves. Models including LSTM, GRU, and Hybrid models of attention-based Conv1D+LSTM and Conv1D+GRU were selected and were trained with 80%, validated with 10% and tested with 10% of the dataset. The proposed model analyses the complicated temporal patterns associated with the occurrence of heatwaves using historical meteorological data from features like temperature, dew point temperature, humidity, wind speed, pressure, rainfall and sunshine. The performance of the models was evaluated using standard techniques, including the Root Mean Squared Error (RMSE), Mean Absolute Error(MAE), and Mean Absolute Percentage Error (MAPE). The attention-based Conv1D+GRU model showed the lowest RMSE and MAPE compared to other models for predicting 7 days ahead maximum temperature and humidity.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

A heatwave is an extended duration of exceptionally hot weather, typically with high temperatures that last longer than the region's historical averages. Heatwaves can have detrimental effects on infrastructure, raise energy costs, and pose major health hazards. They also have negative effects on the economy and ecology. Some of the highest temperatures ever recorded in Bangladesh occurred in 2023. The purpose of this study is to examine several models to determine their predictability and enhance prediction accuracy to aid disaster preparedness and community resilience against extreme heat events in Bangladesh. A heat wave indicates excessive, unseasonably warm temperatures while the heat index helps to quantify the heat being experienced. Since both the heat index and heatwaves are used to evaluate and comprehend the effects of high temperatures on human health and well-being, there is a substantial relationship between the two concepts. By combining humidity and air temperature, the heat index calculates how hot it feels to the human body. In similar ways, heat waves signify intervals of abnormally high temperatures. The assessment of people's perceived temperature during hot weather is the basic concept of both concepts. More extreme heatwave conditions are indicated by higher Heat Index values, which also correlate to higher degrees of discomfort [8].

1.2 Motivation

Heat Waves cause billions of dollars worth of damage and thousands of fatalities every year across the globe. Heatwaves are hitting both developed and developing countries worldwide, and they are not just happening in certain places. They are happening more and more frequently. The terrible effects of the 2022 heatwave, which broke records both in

Bangladesh and globally, highlighted how urgently this problem needs to be addressed. Excessive temperature has created great havoc in Bangladesh in recent years. The temperature remains high during the day, and it doesn't decrease much at night. This prolonged pattern of weather is having a direct impact on the lives, livelihoods, and health of the people in Bangladesh. The extreme weather has increased by 46 percent in the country. This is directly affecting the lives, livelihoods, and health of the people in Bangladesh. Among those aged 65 or older, heat-related deaths have increased by 148 percent. From 2017 to 2021, an additional 1,430 deaths have occurred due to excessive temperatures. This information comes from a report by the Grantham Research Institute on Climate Change and the Environment and the Center for Climate Change Economics and Policy [9]. In the summer of 2021, the temperature increased by 0.49 degrees Celsius, leading to an additional 12 heatwaves and a surge in heat-related deaths, particularly among children. In Bangladesh, almost 21,000 hectares of rice crops were ruined due to heat in 2021 [10]. Only In May, the heatwave had already ruined 141 hectares of rice crops that year. Bangladesh recorded 70 deaths in the July 2022 heatwave only in three days. The local heat trends are influencing global climate change dynamics. Consequently, the quantity, intensity, and duration of heatwaves are on the rise. Simultaneously, the escalating use of biomass combustion is contributing to increased temperature and air pollution, posing a multifaceted risk to public health.

1.3 Socio-economic benefit

Predicting heatwaves in Bangladesh offers a lot of socio-economic benefits, addressing critical challenges posed by extreme heatwave events and enhancing the resilience of communities and the economy. These benefits include:

Protecting Labor Productivity and Economic Output

Dhaka, the capital city of Bangladesh, loses approximately 6 billion USD annually in labor productivity due to heat stress which is equivalent to over 8% of its annual labor output. Each worker loses approximately 10 minutes of work time daily due to extreme heat, totaling 7 billion working hours annually. Accurate heatwave predictions enable industries to implement necessary measures to protect outdoor workers which can reduce productivity losses and minimize economic impacts due to the absence of workers and health-related issues.

Safeguarding Human Health and Reducing Mortality

Heatwaves can cause serious health issues like heat exhaustion, dehydration, heat strokes, and even death, especially among vulnerable populations like the elderly, children, and the poor. Heatwaves have resulted in thousands of deaths in Bangladesh, highlighting the urgent need for preventive measures and proactive healthcare interventions. Predicting heatwaves can enable timely public health interventions, awareness campaigns, and preparedness measures, potentially saving lives and reducing the burden on healthcare systems.

Protecting Agricultural Yields and Ensuring Food Security

Prolonged heatwaves are harmful to food security which causes crop losses and livestock deaths, particularly impacting low-income households reliant on agriculture. It is the agricultural sector, which employs around 40% of Bangladesh's labor force. Heatwave predictions can help farmers take preventive measures, such as adjusting sowing/harvesting schedules or implementing irrigation strategies like implementing heat-resistant crop varieties to protect their yields and livelihoods. Predictive models can aid farmers in mitigating losses and ensuring food security.

Mitigating Infrastructure Risks and Disruptions

During heatwaves, the surface temperature of roads in Dhaka can currently reach up to 60°C (140°F). Such extreme heat can cause melting and damage to road surfaces [11]. During the heatwave in 2023, the Bangladesh Railway authorities had to instruct trains to lower their speed by nearly half of the usual levels. This was done to prevent potential accidents due to the melting or expansion of railway tracks under intense heat. Extreme heat can damage infrastructure like roads, railways, and power grids, leading to disruptions in transportation and energy supply. Heatwave predictions can prompt authorities to take precautions, such as adjusting train schedules or implementing load-shedding measures, to minimize infrastructure damage and maintain essential services.

Informing Urban Planning and Adaptation Strategies

Heatwave predictions can help us to do sustainable urban planning, promoting steps like increasing green spaces, cool roofing, and improving housing conditions to reduce the disproportionate impact on urban poor communities. By addressing vulnerabilities and enhancing adaptive capacities, Bangladesh can build resilience to climate change impacts and rising temperatures. It authorities to establish temporary cooling centers, distribute essential supplies and water, and take other measures to protect vulnerable people, especially those living in urban slums and informal settlements, where access to basic needs is limited. Predicting heatwaves allows tourism authorities to take safety measures and offer alternative activities, minimizing disruptions to the tourism sector and preserving revenue streams.

Overall, accurate heatwave predictions empower decision-makers, communities, and stakeholders to take proactive measures, reducing socio-economic vulnerabilities, and safeguarding essential economic activities and it ensures the well-being and resilience of Bangladesh in the face of increasing extreme heat events.

Chapter 2

Literature Review

We have researched some publications on our proposed topic to have a better knowledge and keep up with the current findings. We discovered two sorts of papers: statistical/environmental and machine learning.

2.1 Related Work

2.1.1 Statistical/Environmental

Samarendra Karmakar et al. [12] researched the heat waves in Bangladesh (1981-2016) where it finds more severe 2014 heat waves in the western region and increasing trends in temperatures $\geq 36°C$, notably in the southwest. Heat waves originate from the west/northwest and extend to central Bangladesh. In 2014, they were influenced by a subtropical high over India and other meteorological factors.

They worked on daily maximum temperature and rainfall from 34 Bangladeshi stations spanning 1981-2018, along with NCEP/NCAR Reanalysis 1 project data, were analyzed. Temperatures $\geq 36°C$ and $\geq 38°C$ for each month from 1981-2016 were computed. Monthly, pre-monsoon, and annual trends and spatial patterns were assessed. In 2014, a year marked by significant heatwaves, extensive tropospheric conditions analysis was conducted for March-May using NCEP/NCAR Reanalysis 1 data. The study aimed to identify the factors driving heatwaves, and the significance of trends was tested using the F-distribution Test.

$$F = \frac{R^2(n-k)}{(k-1)(1-R^2)} \tag{2.1}$$

Where R2 is the coefficient of determination and (n-k) is the degree of freedom.

Bangladesh experiences an increasing trend in annual temperatures ≥ 36°C, with Mongla experiencing the highest increase. Heatwaves are common in the east and center, with distinct activity in the Chittagong Hill Tracts. In 2014, the Jessore-Chuadanga region experienced 54-55 days with temperatures > 36°C, while the Sandwip-M.Court-Feni-Comilla area did not experience them. Higher maximum temperatures advect into Bangladesh from the west and northwest.

M. Bazlur Rashid et al. [13] studied on Global Climate Models (GCMs) where to use a a statistical downscaling approach was applied to produce future climate projections for maximum temperatures during the pre-monsoon season (March-May) in Bangladesh for the 21st century. The analysis used in-situ temperature data from the Bangladesh Meteorological Department, which had recently been digitized. The study compared projected maximum pre-monsoon temperatures in Bangladesh to a reference period of 1981-2010.

In this study, Global Climate Models (GCMs) were used, and a statistical downscaling approach was applied to produce future climate projections for maximum temperatures during the pre-monsoon season (March-May) in Bangladesh for the 21st century. The analysis used in-situ temperature data from the Bangladesh Meteorological Department, which had recently been digitized. The study compared projected maximum pre-monsoon temperatures in Bangladesh to a reference period of 1981-2010. The projections indicated temperature increases in the near future (2021-2050) and the far future (2071-2100) under different emission scenarios: 1. Under the RCP8.5 scenario, temperatures are projected to increase by 0.7°C in the near future and 2.2°C in the far future. 2. Under the RCP4.5 scenario, temperatures are projected to increase by 0.7°C in the near future and 1.2°C in the far future. 3. Under the RCP2.6 scenario, temperatures are projected to increase by 0.7°C in the near future and 0.8°C in the far future.

The empirical-statistical downscaling approach used in this study involved deriving statistical relationships between the mean daily maximum temperature over the pre-monsoon season from station-based observations in the period 1981–2019 and the large-scale mean temperature patterns as represented by the common EOFs of the ERA5 reanalysis and GCM simulations.

The GCM data (one simulation at a time, using the full period of available data 1850–2100) was combined with ERA5 reanalysis data for the period 1979–2019 along the time axis. EOF analysis was subsequently applied to the combined dataset.

PCA decomposes observed data into spatial patterns, Principal Components (PCs), and eigenvalues.The first two leading PCs, accounting for a significant portion of observed variability (94% (80% for PC1 and 14% for PC2)), are used as predictors.This procedure separated the data into common spatial patterns and eigenvalues shared between reanalysis and GCM data, with individual principal components (PCs) reflecting temporal variations, one part associated with reanalysis data and the other with GCM data. The PCs associated with reanalysis data are used to calibrate statistical models.

Step-by-step multiple regression to determine model parameters, utilizing data from the calibration period of 1981-2019, where only the predictor PCs associated with reanalysis were included in the regression.The statistical models are then applied to predictor PCs associated with GCM data to obtain future projections of predictor PCs, Then they constructed future estimates of local maximum temperatures by combining the projections of the predictand PCs with the corresponding spatial patterns and eigenvalues.The entire process of common EOF analysis and regression is repeated for each GCM simulation, allowing for efficient application to multiple simulations. Validation includes a five-fold cross-validation approach to assess model skill and a target plot to evaluate the representation of past trends and interannual variability.Past climate data (1981-2019) revealed a significant increase in maximum temperatures during the pre-monsoon season, with the most substantial changes observed in the eastern region.Statistical downscaling models were successfully validated with PC1 (a strong correlation with reanalysis) and PC2(a significant correlation). Projections of substantial temperature increases for both the near and far future under different emission scenarios.The strongest warming was projected under the high-emission scenario (RCP8.5).

2.1.2 Machine Learning

Siqin Wang et al. [14] combined exposure to heat and air pollution affects physical and mental health in Australia. It finds that social and built environmental factors are more important than temperature and air quality, and that the relationship between these variables and health follows a V-shape. The importance of heat and air pollution is most obvious in urban areas and industrial zones, and the study draws policy implications to minimize healthcare access and service distribution differences. The study presents insights into how combined heat and air pollution exposure affects health outcomes, using a spatially explicit method that integrates machine learning techniques and spatial algorithms. The evidence is useful for place-based health planning, risk profiling, and designating regional strategies in global warming mitigation in the southern hemisphere. The study calls for joint efforts from multiple disciplines and sectors to advance Australia's position in the climate-health research agenda and make the country climate-resilient.

Landscape surface temperature data may have discrepancies to air temperature, which affects human sensation of heat. Further calibration is needed for accurate heat measurement. Future studies should use raster-based regression modeling and consider additional confounders to better understand the environmental relationship and the interactions among variables including individual-specific factors (e.g., frequency of exercises and living style), regional factors (e.g., regional weather conditions, urban morphological and typological factors (e.g., elevation and slope). Despite the reasonably high performance of the global and local RFR models, our analyses would not be able to reveal the causality between heat & air quality and health and therefore the modeling results should be interpreted with great caution. longitudinal cohort studies are needed to track the long-term causal impact of the environment on human health, though it is challenging and difficult to implement with the same cohorts consistently over time. The use of global and local machine learning models in environmental research is limited due to technical drawbacks such as the Blackbox issue, less interpretability and generalizability, and difficulty in replication compared to frequentist statistical models. The article highlights the complexity of the relationship between environmental factors and human health. Machine learning and spatial algorithms are employed to provide nuanced insights and inform health planning and risk profiling.

Seyed Babak Haji Seyed Asadollah et al . [15] introduced an approach for predicting Heatwave Days (HWDs) in Iran using Machine Learning (ML) models. This study addresses rising temperature extremes, notably heatwaves, due to climate change. It employs machine learning models, including decision trees, random forests, and a novel hybrid technique (ABR-DT), to forecast Iran's annual heatwave days. Data from Princeton Meteorological Forcing and National Centers for Environmental Prediction reanalysis are used. ABR-DT outperforms other models with a correlation coefficient of 0.860 and a mean absolute error of 6.929, using specific humidity and wind components as predictors. The approach proves effective in predicting heatwave days across Iran's diverse climate regions. Decision Tree (DT), Random forests, Ada boost regression—decision tree (ABR-DT) machine learning algorithms were used. By these methods, it show a higher correlation (up to 0.65) with variables at the 850 and 1000 hPa, representing higher consistency with HWD than those at low-pressure levels. It showed an average correlation of 0.62 and 0.54 for G7 and G4, while the wind components (uwnd and vwnd) showed the lowest correlation for all climate regions. The results show that ML models are effective in forecasting HWDs, with ABR-DT proving the most reliable. Specific humidity and wind components are found to influence HWDs in Iran. This methodology can potentially be applied to develop heatwave forecasting models in other regions worldwide. Future research may involve comparing ML model results with dynamical models and exploring additional ML techniques for improved forecasting.

Peiyuan Li et al. [16] have proposed a heatwave prediction algorithm based on GNN(Graph Neural Network) to provide warnings of sudden regional summer heat waves with high accuracy and lower costs of computation and data collection. They have selected 91 stations in the Contiguous United States (CONUS) with consecutive daily recordings in the 15 years from 1 January 2006 to 31 December 2020 (5,479 days). The identification of HW(heatwave) events is measured by the maximum temperature for each station that is when the maximum temperature is above the 90th percentile of each maximum temperature for each day and stays at least 3 consecutive days from May to September. The moving window of daily Tmax is 15 days. They have pre-labeled the days that HW occurs with 1 and 0 otherwise. They have formulated the network by treating weather stations as nodes and generating edges based on the correlation coefficients (CC) between two nodes calculated from the daily maximum temperature and precipitation time series over 15 years. They have set the cut-off thresholds to be the 0.75 quantiles of the CCs calculated from all nodes. This treatment ensures every node has at least one and an average of 10 neighbour nodes.They have selected the first 13 years of data (2006–2018) for model training and the last 2 years (2019 and 2020) for model validation.The proposed method in the thesis paper uses a Graph Neural Network (GNN) model with encoder-processor-decoder architecture used for heatwave prediction. Here binary classification was used to forecast heatwave occurrences from input features. Soft F1-score is used to reduce model bias while training the model. The GNN model utilizes an attention-based graph neural network with a Graph Attention Layer (GAL) for information exchange between weather stations.

The GNN model predicts the HW based on the non-linear relations it learns from the historical data. They have conducted the sensitivity test over the 14 input features using the nominal model.The relative sensitivity is quantified by the normalized sensitivity score (SS). The SS values also vary from month to month as the GNN model captures seasonal dynamics from the training data. The GNN model achieved a 94.1% accuracy, with 58.5% recall and 62.5% precision for heatwave prediction.It successfully captured spatial patterns that is it was accurate to identify and represent how heatwave events are distributed across different geographical locations. The model faced challenges to identify specific locations of heat wave events. It is not as effective in capturing how these patterns change over many years.

Minsoo Park et al. [17] developed a random forest model for the weekly prediction of heat-related damages on the basis of four years (2015–2018) of statistical,meteorological, and floating population data from South Korea. The model was evaluated through comparisons with other traditional regression models in terms of mean absolute error, root mean squared error, root mean squared logarithmic error, and coefficient of determination (R2). In a comparative analysis with observed values, the proposed model showed an R2 value of 0.804. They have selected the following variables: temperature, humidity, wind speed, number of vulnerable occupational groups, insurance premiums per person, personal income per person, floating population, and registered population of residents.The resolution of the entire dataset was unified through considering the data properties of the features and targets; the temporal resolution was set to 1 week, and the range resolution was set on the basis of the South Korean administrative divisions.80% were used as learning data and the remaining 20% as test data. All variables were confirmed using the Boruta algorithm.

They have used hyper-parameter optimization. The hyperparameters were determined with

181 decision trees and 46 tree depth.The RF model was trained on the basis of the determined hyperparameters, and test data were applied to the regression model.The RF model is compared with linear regression, decision tree, and support vector machine (SVM) models.The best values for all the considered metrics, including MAE (3.816), RMSE (8.655), RMSLE (0.645), and R2 (0.803), were obtained for the RF model.

A temperature prediction model that covers the city of Agadir, Morocco proposed by Anas Kabbori et al. [18]. They have used time Series Time-Delay Neural Network, a type of dynamic neural network.They have considered the availability of data for previous month to predict the temperature values for the next month. Their proposed model has 3 layers- input layer,one hidden layer and output layer. The activation function used with the hidden layer neurons is the hyperbolic tangent function, and identity function for the output neuron. The network training used Backpropagation Through Time (BPTT) in epoch wise mode. The training uses LevenbergMarquardt backpropagation algorithm.

The training data was gathered from the Meteorological station located in Agadir Airport and the data represents temperature values through the station on a 30 min basis 24 hours and 7 days a week. The data set for the first 28 days are used to train the proposed network, while the data set for the last two days is used to test the trained network.

The performance of the network has been tested with MSE, training regression, Error autocorrelation, time-series respond and error histogram. Taking the last 48 hours of June 2019 as example gave them a Mean Squared Error (MSE), between the actual response and the target for the testing data of a value of 0.4149.The computed value of The correlation coefficient using the testing data is about 0.99. They have found their best training performance at epoch 1000 which is 0.17215.

Najeebullah Khan et al. [19] shown a comparison between Quantile Regression Forests (QRF) and Random Forest (RF) for the prediction of heat waves in Pakistan. According to(Khan et al., 2018c), the heat wave in Pakistan is defined as daily maximum temperature above 95th percentile of maximum temperature of the base year (1971–2000) for at least 5 consecutive days (Khan et al., 2018c).The main objective of the present study is to explore the capability of QRF to predict heat waves with various levels of accuracy.

The Princeton's Global Meteorological Forcing (PGF) daily maximum temperature data for the period 1948–2010 was used to reconstruct the historical heat waves and the National Centers for Environmental Prediction (NCEP) reanalysis synoptic climate variables were used for the selection appropriate set of predictors for the forecasting of heat waves with 1–10 days lead times using QRF.The daily gridded maximum temperature of PGF was used due to its higher spatial resolution of $0.25° \times 0.25°$. The QRF and RF models were developed to predict daily maximum temperature of Pakistan.The models were calibrated for the period 1948–1991 and then validated for the period 1992–2010.Separate models were developed using separate set of predictors for the months of May, June and July and the performance of the models in prediction of the trigger and the departure dates of heat waves for different lead times ranging from 1 to 10 day(s) were estimated.A domain consisting of 357 grid points covering the region between latitudes 40- 90E and longitude 10-50 N was selected, considering that it is sufficient to cover the influence of circulation patterns responsible for daily temperature of Pakistan.

Comparatively, a much higher rate of failure was noticed for RF than QRF in prediction with different lead times. The results also revealed that wind and relative humidity are the major factors that define heat waves in Pakistan. The QRF model was observed to capture more than 70% of the triggering date of heat waves with an accuracy of ±3 days up to 10-day lead time for the month of June which is the most potential month of the occurrence of heat waves in Pakistan. It is also found to predict the triggering date of heat waves for 60% cases with the same accuracy in the month of May and 50% in the month of July.

Saqiba Juna et al. [20] used LSTM and were successful in predicting heatwaves in Pakistan and suggested the importance of specific weather variables in anticipating extreme temperature events. This study focused on the challenge of heat waves caused by global warming, with a specific emphasis on Pakistan. To tackle this issue, the researchers use Long Short-Term Memory (LSTM) neural networks, a type of deep learning technology, to predict the maximum temperatures during heatwaves. They collected and preprocessed data from the Karachi region for this purpose. The LSTM model's performance is evaluated using several metrics, including Mean Square Error (MSE) and Root Mean Square Error (RMSE). Additionally, the study introduces a novel percentile-based threshold approach that effectively predicts heatwave events, contributing to our understanding and mitigation of the impacts of rising temperatures due to global warming. Here they employed a Long Short-Term Memory (LSTM) model, a type of sequential model, to forecast heatwaves in the Karachi region. The methodology followed a structured approach, including data collection, preprocessing, LSTM model development, and post-processing with a regional threshold value for heatwaves. Temperature data was collected from the World Weather Online website, spanning from January 2010 to December 2020. Preprocessing involved addressing noise, missing data, and outliers, followed by feature normalisation. The LSTM model was specifically designed to analyse time series temperature data for heatwave prediction. It utilised various hyperparameters, including the number of neurons, hidden layers, epochs, activation function, optimizer, dropout, and batch size. The final step involved post-processing by applying a regional threshold value to identify and characterise heat waves. The performance of the LSTM model was evaluated using metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). The data was divided into training and validation sets. The LSTM model outperformed other deep learning algorithms. Post-processing involved applying regional heatwave thresholds, determined by calculating the 95th percentile of maximum temperature. In Karachi, a threshold of 40 degrees Celsius was set.

The use of machine learning to predict heatwave severity by training models on historical temperature data proposed by Aniruddha Mane et al. [21] Machine learning models are used to offer more accurate predictions than conventional Statistical forecasting methods, with a focus on the highest-performing model among five regression models. The study primarily focuses on Telangana, a region particularly susceptible to heatwaves, with some districts in severe or critical condition.

After web scraping they collected data which includes parameters like temperature, humidity, and wind velocity. Data preprocessing involves organising the data into year-by-year monthly meteorological records stored in CSV format files. Data cleaning is performed to eliminate errors and irrelevant information.Then they completed the model training part using supervised machine learning models including linear regression, random forest regressor, lasso regressor, support vector machine, and neural network regressor. In the model training process 80% of the total dataset was used as a training sample for each prediction model. Testing is carried out to evaluate model performance, with 20% of the data used as testing samples. Model evaluation includes the calculation of accuracy and loss using metrics like R2 score, MSE, and MAE. Multiple models are compared, and a user interface is developed using Streamlit to make heat index predictions based on input parameters like air temperature and dew point temperature. The Random Forest Regressor algorithm is used for predicting the heat index in the interactive web application.

The comparison of accuracy and loss parameters revealed that Random Forest Regressor outperformed other models, achieving an R2 score of 99.9954, an MAE of 0.0146, and an MSE of 0.01206. The findings underscore the effectiveness of Random Forest Regression for handling high-dimensional, non-linear data and reducing variance. The study highlights the potential of AI in heatwave prediction, offering valuable insights for applications in mitigating heatwave impacts on various sectors. Future research could explore broader geographical coverage and real-time data integration to enhance the model's utility.

Clyde Shelton Bangera et al. [22] predicted floods and heat waves in Mangalore involves analyzing weather data using a combination of techniques. They focused on using machine learning techniques to predict natural calamities. Specially Flood & Heat waves in Mangalore, India.It introduces an approach that combines weighted moving averages and K-Nearest Neighbors for flood prediction and utilizes anomaly detection for identifying heat waves. The research seeks to enhance preparedness and response systems using cutting-edge machine learning methods. To predict floods, temperature, humidity, and pressure data are collected and processed using a weighted moving average (WMA) method with a 7-day window. The WMA results are then fed into K-Nearest Neighbors (K-NN) classifiers for rainfall prediction. The value of K for K-NN is determined experimentally, and the process is repeated for consecutive days to forecast floods. The accuracy of weather parameter predictions using WMA ranges from 85% to 93%, with an error of around 4-5%. Incase of predicting Heat wave the predicted temperature is compared with the average temperature of the region, which is calculated using the monthly average temperatures of the previous three years. If the predicted temperature exceeds a certain threshold above the average, a heat wave is detected. This methodology combines data analysis and machine learning to predict natural calamities in Mangalore.

Zao Zhang et al. [23] conducted a comparative study of CNN and RNN in forecasting. They worked on a deep-learning model for precise temperature forecasting. It stands out from traditional meteorological approaches by adapting to various geographic regions, proving valuable where there are no well-established meteorological models. The accuracy of the model is improved by its independent learning of time and space correlations from historical data. CNNs excel at two-dimensional data processing, whereas RNNs specialize in sequential data. By effectively managing two-dimensional data with spatial and temporal correlations, CRNN combines the strengths of both. CNN extracts embedded information and processes spatial correlations, while RNN handles temporal correlations, making it appropriate for tasks requiring both time and space-dependent features.

2.2 Summary Table

Table 2.1: Summary Table of Literature Review

SL	Title	Methodology	Results
1	Unpacking the inter- and intra urban differences of the association between health and exposure to heat and air quality in Australia using global and local machine learning models	Principal component analysis, Random forest regression, Accumulated local effects analysis.	Machine learning and spatial algorithms, reveals the connection between environmental factors and human health, which offers insights for informed health planning and risk profiling.

Table 2.1 continued from previous page

SL	Title	Methodology	Results
2	Prediction of heat waves using meteorological variables in diverse regions of Iran with advanced machine learning models	Decision Tree (DT), Random forests, Ada boost regression decision tree (ABR-DT) machine learning algorithms were used.	The results show that ML models are effective in forecasting HWDs, with ABR-DT proving the most reliable. Specific humidity and wind components are found to influence HWDs in Iran.
3	ON THE HEAT WAVES IN BANGLADESH, THEIR TRENDS AND ASSOCIATED LARGE SCALE TROPOSPHERIC CONDITIONS	F-distribution Test.	The study aimed to identify the factors driving heatwaves.
4	Regional Heatwave Prediction Using Graph Neural Network and Weather Station Data	Graph Neural Network (GNN).	94.1% accuracy, 58.5% recall and 62.5% precision.
5	Heatwave Damage Prediction Using Random Forest Model in Korea	Random Forest(RF), Linear Regression, Decision tree, and Support Vector Machine (SVM).	MAE (3.816) RMSE (8.655) RMSLE (0.645) and R2 (0.803)
6	Temperature Prediction using Time Series Time Delay Neural Networks	Time Series Time-Delay Neural Networks, LevenbergMarquardt backpropagation algorithm BackPropagation Through Time in epoch wise mode.	MSE(0.4149). The correlation coefficient was 0.99.

Table 2.1 continued from previous page

SL	Title	Methodology	Results
7	Prediction of heat waves in Pakistan using quantile regression forests	Comparison between Quantile Regression Forest(QRF) and Random Forest(RF).	Accuracy 70% (June) 60% (May), 50% (July).
8	Climate change projections of maximum temperature in the pre-monsoon season in Bangladesh using statistical downscaling of global climate models.	The empirical-statistical downscaling approach, Principal Component Analysis(PCA), ERA5 Reanalysis, GCM simulation.	Statistical downscaling models were successfully validated with PC1 and PC2.The strongest warming was under the high emission scenario (RCP8.5).
9	Regional Heatwave Prediction Using deep learning based Recurrent Neural Network.	Long Short Term Memory(LSTM).	93%, 87%, and 85% accuracy for temperature, humidity, and pressure, respectively.
10	Flood and Heat Wave Prediction using Weighted Moving Average, Anomaly Detection, and K-Nearest Neighbours for the City Of Mangalore	Weighted Moving Average(WMA)and K-Nearest Neighbors(KNN).	90% of accuracy.
11	Temperature Forecasting via Convolutional Recurrent Neural Networks Based on Time-Series Data	Convolutional Recurrent Neural Networks(CRNN).	RMSE(1.697), and high accuracies for prediction errors below 1, 2, and 3°C which are 0.689, 0.830, and 0.914 respectively.

Table 2.1 continued from previous page

SL	Title	Methodology	Results
12	Artificial Intelligence Based Heatwave Intensity Prediction Model	Random forest, Artificial Neural Network(ANN), SVM, LASSO regressor, and Linear Regressor.	R2 (99.9954), MAE(0.0146), MSE(0.01206).

2.3 Research Gap

The study [17] has several limitations too. Initially, the temporal resolution of the predicted values is relatively coarse-grained. Next, during the experiment, data obtained across the country were input into one learning algorithm, and regional differences between administrative districts were not considered.

Another study [22] has highlighted the effectiveness of K-Nearest Neighbors (KNN) for accurate rainfall prediction and, subsequently, flood detection, as it relies on lightweight calculations.

As we discovered in Bangladesh, some research is based solely on environmental or statistical models, but there is not much work applying models based on machine learning for heatwave prediction. We want to examine which deep learning model offers better output for heat prediction with our country's data because there are no significant machine learning-based work has been done on heat waves in Bangladesh.

Chapter 3

Background Study

Certain Deep learning models are used for enhancing the predictability of heatwaves in Bangladesh so that a comparative study on those deep learning models can be built depending on their accuracy. So before diving into that these models should be described first.

3.1 Long Short-Term Memory

Long Short-Term Memory (LSTM) is a recurrent neural network architecture designed by Sepp Hochreiter and Jürgen Schmidhuber in 1997.

The LSTM architecture consists of the memory unit. The LSTM memory unit consists of four feedforward neural networks also known as gates. They are the forget gate, the input gate, the candidate gate, and the output gate.

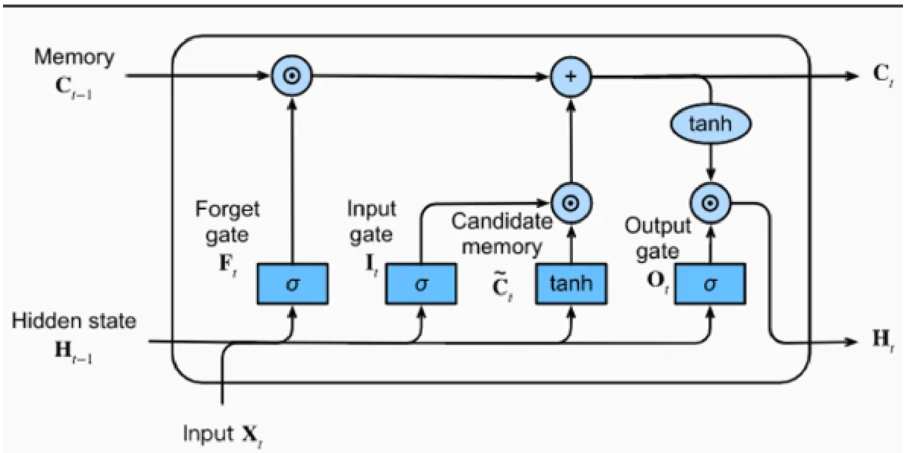


Figure 3.1: The architecture of an LSTM Unit [2]

Among them, the forget, input and output gates are responsible for memory management.

All three gates use the sigmoid function as the activation function in the output layer and this sigmoid function is responsible for the flow of information. With the help of these three gates, the LSTM unit maps an input to an output sequence using these equations.

$$i_t = \sigma(W_i \cdot [h_{t-1}, X_t] + b_i) \tag{3.1}$$

$$\widetilde{C}_t = \tanh(W_c \cdot [h_{t-1}, X_t] + b_c) \tag{3.2}$$

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, X_t] + b_f\right) \tag{3.3}$$

$$C_t = i_t * \widetilde{C}_{t-1} + f_t * C_{t-1} \tag{3.4}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, X_t] + b_o) \tag{3.5}$$

$$h_t = o_t * \tanh(C_t) \tag{3.6}$$

In the LSTM unit, the forget gate is responsible for deleting unwanted information from memory. Input gates conditionally insert information from the input sequence to update the memory. The output gate is responsible for giving the output after analyzing information in memory [24] [2].

3.2 Gated Reccurent Unit

The Gated Recurrent Unit (GRU) is a type of recurrent neural network (RNN) that offers improvements over traditional RNNs and Long Short-Term Memory (LSTM) networks in sequence modeling tasks.GRU utilizes gates to regulate the flow of information within the network like LSTM,. But it has a simpler architecture than LSTM. GRU has only two gates: the Reset Gate and the Update Gate. It does not have a separate cell state, unlike LSTM, which contributes to faster training due to its simpler structure.

The architecture of GRU isn't very complex. At every timestamp t, it takes an input X_t with the hidden state H_{t-1} from t-1 which is the previous timestamp. After that, it outputs a new hidden state H_t and it is passed to the next timestamp. The short-term memory of the network is controlled by the Reset gate, represented by the following equation:

$$r_t = \sigma(x_t \cdot U_r + H_{t-1} \cdot W_r) \tag{3.7}$$

The sigmoid function causes the value of r_t to range from 0 to 1. Here weight matrices for the reset gate are U_r and W_r. The Update Gate controls long-term memory and is defined by

$$u_t = \sigma(x_t \cdot U_u + H_{t-1} \cdot W_u) \tag{3.8}$$

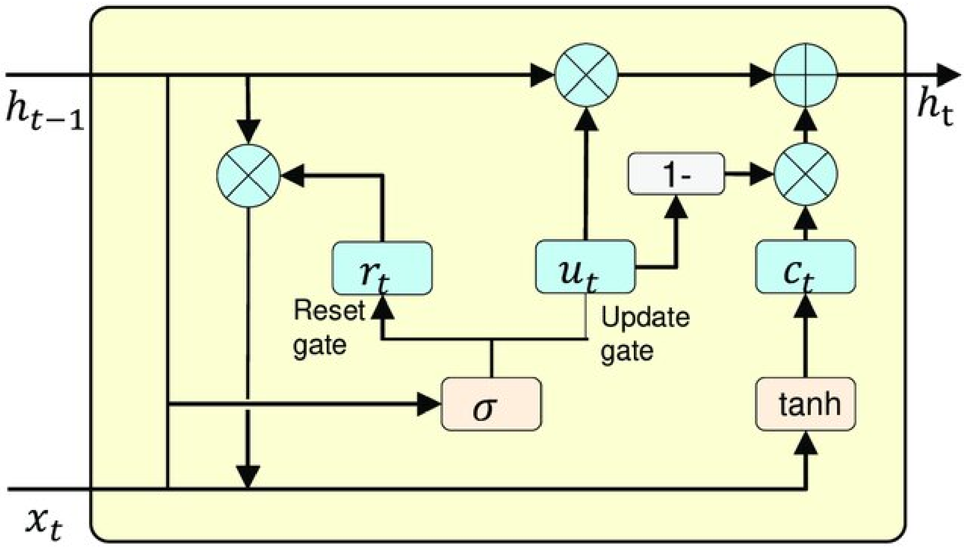


Figure 3.2: Basic Architecture of GRU [3]

Here weight matrices for the update gate are U_u and W_u.

GRU works in fascinating ways. It generates a candidate hidden state, which is a preliminary version of the hidden state. It calculates as:

$$\tilde{H}_t = \tanh(x_t \cdot U_g + (r_t \odot H_{t-1}) \cdot W_g) \tag{3.9}$$

It takes input and the hidden state from the previous timestamp t-1 then it is multiplied by the output of the reset gate r_t. After that it passes this entire information to the tanh function, the resultant value is the candidate's hidden state. When the value of r_t is equal to 1 then it indicates that all the information from the previous hidden state H_{t-1} is being considered. And if the value of r_t is 0 that means the information from the previous hidden state is not considered. After calculating the candidate state, the current hidden state (H_t) is generated using the Update gate. Unlike LSTM's separate gates, GRU uses a single Update gate to control the historical information (H_{t-1}) and the new candidate state information.

$$\mathrm{H}_t = \mathrm{u}_t \odot \mathrm{H}_{t-1} + (1 - \mathrm{u}_t) \odot \tilde{H}_t \tag{3.10}$$

If the Update gate (u_t) value is 0, the new hidden state (H_t) consists mostly of the candidate state, ignores the previous hidden state (H_{t-1}). On the other hand, if H_t is close to 1, H_t relies entirely on H_{t-1}, disregarding the candidate state. Thus, u_t's value between 0 and 1 controls the balance between historical and new information in the hidden state.

GRU addresses gradient vanishing and exploding issues in RNNs with a more simplified design compared to LSTMs, it offers higher computational efficiency and effectively captures long-range dependencies in sequential data

3.3 Convolutional Neural Network

A convolutional neural network is one of the most special kinds of neural networks that has come up with breakthrough results on various pattern recognition problems. Basically, CNN is a form of neural network which works with identical copies of the same neuron. This helps the network to have a large number of neurons, creating complex computational models while keeping the number of actual parameters relatively small and manageable.

A CNN typically has three layers: a convolutional layer, a pooling layer, and a fully connected layer.

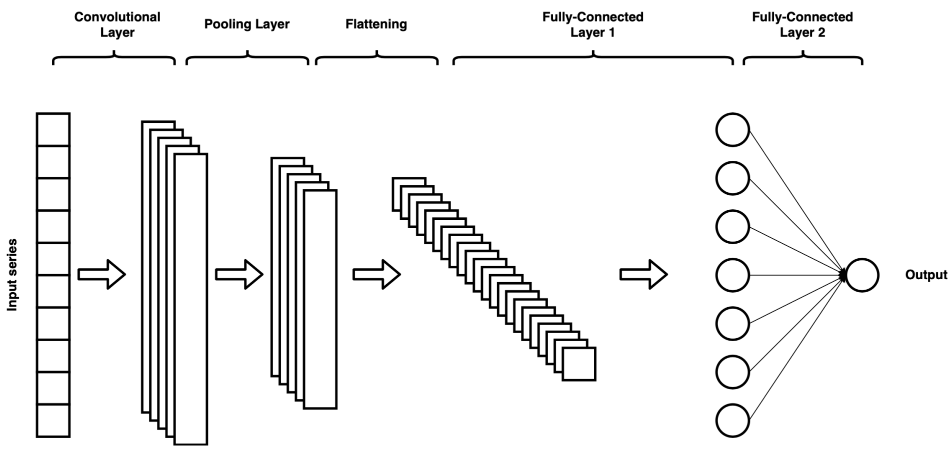


Figure 3.3: Architecture of a convolutional neural network [4]

But in our work, we will be using 1 dimensional CNN also known as Conv1D. A 1D Convolutional Neural Network (Conv1D) is mostly used for sequence data, such as time series or text.

Convolutional Layers: Convolutional layers in a Conv1D network are generally designed to capture relationships and patterns in sequential data.

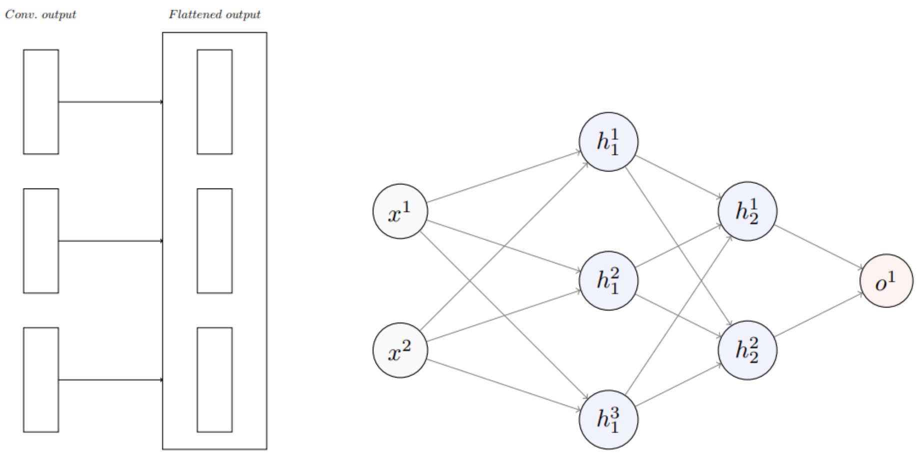
The convolution operation in 1D involves sliding a filter (kernel) along the sequence and computing the dot product at each step.

Let x be the input sequence, w be the filter, b be the bias, and * denote convolution:

$$(x \cdot w + b)_i = \sum_{j=0}^{M-1} x_{i+j} \cdot w_j + b \tag{3.11}$$

Here, M is the size of the filter.

After convolution, the result goes through an activation function, commonly ReLU:



(a) Flatten output (b) Fully connected network

Figure 3.4: Flatten output and fully connected network [5]

$$\text{ReLU}(x) = \max(0, x) \tag{3.12}$$

Pooling Layers: Pooling layers in Conv1D downsample the data to reduce its dimensionality.

Max-Pooling Operation: Max-pooling in 1D takes the maximum value from a group of values in the input sequence:

$$\text{Max-Pooling}(x) = \max(x) \tag{3.13}$$

Fully Connected Layers: Predictions are made using fully connected layers based on the learned features.

Flattening: Before fully connected layers, the data is mostly flattened from a 1D sequence to a 1D vector.

Fully Connected Operation: The fully connected operation involves computing the weighted sum of input values and passing the result through an activation function:

$$\text{Output} = \text{Activation}\left(\sum_i (Weight_i \times Input_i) + \text{Bias}\right) \tag{3.14}$$

This structure allows Conv1D networks to naturally learn features and patterns in sequential data to make them effective for tasks like time series analysis [5] [25].

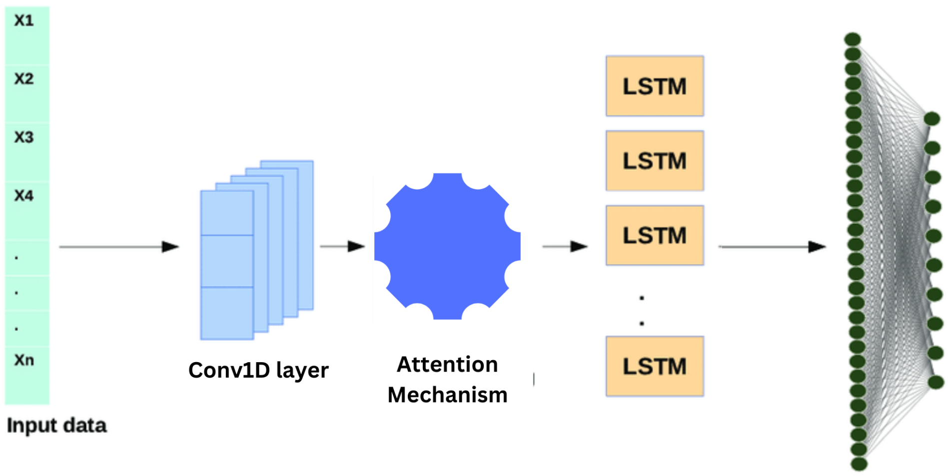


Figure 3.5: Basic Structure of Conv1D+LSTM with attention mechanism [6]

3.4 Attention-Based Hybrid Models

The Conv1D+LSTM [26] hybrid model combines the strengths of Convolutional Neural Networks (Conv1D) and Long Short-Term Memory (LSTM) networks. To effectively capture both local patterns and long-term dependencies in sequential data this hybrid architecture is designed. As we know, for time series forecasting understanding both short-term fluctuations and long-term trends is essential. This architecture is particularly well-suited for tasks like that. Based on the characteristics we also know that Conv1D is efficient at extracting features and creating perfect representations of time series data, and the LSTM is known for its skills in finding long-term temporal dependencies of time series data. The Conv1D+LSTM hybrid model, enhanced with an attention mechanism, combines the strengths of Convolutional Neural Networks (Conv1D), Long Short-Term Memory (LSTM) networks, and selective attention mechanisms.

The hybrid model starts with an input layer, which receives the input sequential data. This input data is then passed through one or more Conv1D layers, where local patterns in the data are identified. Each Conv1D layer consists of multiple filters, each extracting specific features from the input sequence. The choice of kernel size and activation function in these Conv1D layers influences the model's ability to capture relevant patterns.

After the Conv1D layers, an attention mechanism is introduced. This mechanism selectively focuses on crucial segments of the input sequence and dynamically allocates weights to different parts based on their significance. By attending to relevant information, the model can effectively prioritize prominent features, enhancing its forecasting accuracy.

After attention processing, the data goes into one or more LSTM layers. Here, the LSTM units meticulously process the sequential input, adeptly retaining information across extended periods. This sequential processing capability empowers the model to capture intricate temporal dependencies, thereby making accurate forecasting of future time steps.

At the end of the model, there is a layer that makes the final predictions based on all the information it has learned from all the sequential data. This layer usually consists of a dense layer, which is like a filter, with a specific way of deciding the final prediction. This decision-making process is perfected to match the type of problem that needs to be solved.

Conv1D layers smooth out the input time-series data and eliminate the need for additional preprocessing steps like adding rolling mean or rolling standard deviation values to the input features. LSTM networks can effectively model problems with multiple input variables, thus making them suitable for multivariate time-series forecasting tasks. The 3D input vector required by LSTM accommodates the multivariate nature of the data.

The Conv1D-LSTM model supports various sequence-to-sequence LSTM architectures, including many-to-one and many-to-many models. This flexibility allows for a variety of forecasting scenarios, such as forecasting at the current timestep or predicting multiple future time steps simultaneously [27]. The addition of an attention mechanism helps the Conv1D+LSTM architecture by enabling the model to focus on the most important aspects of the input sequence. This targeted attention mechanism enhances and improves the model's interpretability and forecasting accuracy, making it perfectly skilled for a variety of time series forecasting tasks [28].

The Attention-Based Conv1D+GRU model, similar to its LSTM counterpart, combines Convolutional Neural Networks (Conv1D) with Gated Recurrent Unit (GRU) networks, enhanced with an attention mechanism. This hybrid architecture effectively captures both local patterns and long-term dependencies in time-series data. Starting with an input layer, the model utilizes Conv1D layers for feature extraction, followed by an attention mechanism to focus on crucial segments of the input sequence. The GRU layers then process the sequential input, adeptly retaining information across extended periods. The addition of the attention mechanism enhances the model's interpretability and forecasting accuracy, making it suitable for various time series forecasting tasks [29].

3.5 Weather Research and Forecasting (WRF) Model

The Weather Research and Forecasting (WRF) Model [30] is a state-of-the-art mesoscale numerical weather prediction system that is designed for both atmospheric research and operational forecasting applications. It has a wide range of meteorological applications across scales ranging from tens of meters to thousands of kilometers. The development of WRF began in the late 1990s as a collaborative effort between several institutions, including the National Center for Atmospheric Research (NCAR), the National Oceanic and Atmospheric Administration (NOAA), the U.S. Air Force, the Naval Research Laboratory, the University of Oklahoma, and the Federal Aviation Administration (FAA). The WRF Model has two dynamic cores, they are the Advanced Research WRF (ARW) and the Nonhydrostatic Mesoscale Model (NMM). These cores compute the atmospheric governing equations. The model also includes a data assimilation system and a software architecture that supports parallel computation and system extensibility.

The methodology [31] used by the WRF Model includes simulating atmospheric conditions based on either real data (observations and analyses) or ideal conditions. The dynamical cores solve the governing equations, which take various atmospheric processes into account, such as precipitation, cloud formation, and atmospheric dynamics. The model incorporates advanced physics, numerics, and data assimilation techniques contributed by the research community. The WRF Model offers two main dynamical solvers for computing the atmospheric governing equations: ARW (Advanced Research WRF) and NMM (Nonhydrostatic Mesoscale Model). The ARW solver is supported by the NCAR Mesoscale and Microscale Meteorology Laboratory, while the NMM solver is supported by the Developmental Testbed Center (DTC). In the context of Bangladesh, the WRF Model is used to forecast the weather by simulating atmospheric conditions over the region. It can analyze real-time observations and analyses from various sources to initialize the simulations. The dynamical cores then solve the governing equations, considering complex atmospheric processes, to provide high-resolution forecasts for Bangladesh from the GFS data. The Global Forecast System (GFS) data refers to numerical weather prediction output produced by another model called the GFS model, which is operated by the National Centers for Environmental Prediction (NCEP), a division of the National Oceanic and Atmospheric Administration (NOAA) in the United States. The GFS model is a global numerical weather prediction system that mainly simulates atmospheric conditions across the entire globe. GFS data provides valuable information about the initial state of the atmosphere, including temperature, humidity, and wind fields, which serve as the starting point for WRF simulations. By assimilating GFS data into its initial conditions, WRF can initialize its simulations with a more accurate representation of the current atmospheric state of local areas of Bangladesh. The WRF Model has a large worldwide community of users (over 30,000 registered users in over 150 countries), and NCAR provides regular workshops and tutorials for its users. Worldwide WRF Model carries a lot of value in the field of weather analysis.

3.6 Deep Learning Methods for Weather Forecasting

LSTM (Long Short-Term Memory)

LSTMs are well-suited for capturing long-term dependencies in sequential data. In climate data, where the occurrence of heatwaves is likely influenced by historical weather patterns, LSTMs can effectively learn and model the temporal dependencies between different time steps.

The memory cell in an LSTM allows the model to selectively remember or forget information from past time steps. This is particularly beneficial when dealing with climate data, as certain weather conditions may have a lasting impact on the likelihood of heatwaves. The memory cell helps the model retain relevant information over extended periods.

LSTMs support sequence-to-sequence learning, making them suitable for predicting sequences of events. In the context of heatwave prediction, where the occurrence and persistence of extreme temperatures form a sequence, LSTM's sequence modelling capabilities are advantageous.

LSTMs have shown success in various time-series forecasting tasks, including weather prediction. Their architecture allows them to capture and learn complex temporal patterns, which is crucial for accurately predicting heatwaves that exhibit temporal dependencies.

LSTMs have been widely used in climate modelling and weather forecasting due to their ability to capture intricate patterns in time-series data. This previous success makes them a natural choice for predicting heatwaves in Bangladesh.

LSTMs can learn both short-term and long-term patterns, making them versatile in capturing various types of dependencies in the data. This flexibility is important in climate data analysis, where the factors influencing heatwaves may have different time scales.

As a result of their proven performance in climate modelling and weather forecasting, as well as their capacity to manage temporal dependencies and accurately represent long-term patterns, sequence-to-sequence learning is supported by LSTM models. The goal of applying LSTMs to your climate dataset is to make use of these advantages in order to forecast heatwaves in Bangladesh with accuracy and dependability.

GRU(Gated Recurrent Unit)

GRU (Gated Recurrent Unit) is a significant tool in predicting heatwaves from time series weather data due to its adeptness in capturing complex temporal dependencies. It can handle long-range sequences, and efficiently process multivariate data. Heatwaves are characterized by prolonged periods of exceptionally high temperatures, which need models capable of discerning subtle patterns and trends within historical weather data. GRUs are good at keeping important information for a long time, which helps them figure out when heatwaves might happen accurately. Also, GRUs have a simple design that makes it easier and faster to train them and use them to look at lots of weather data quickly, especially when we need to do it in real-time situations. The ability of GRUs to handle multivariate inputs also enhances their ability in heatwave prediction, allowing for the integration of various meteorological variables that contribute to heatwave formation. By harnessing the strengths of GRUs, meteorologists and climate scientists can develop robust heatwave prediction tools, empowering proactive planning and response strategies to mitigate the impacts of extreme heat events.

Conv1D (1D Convolutional Neural Network)

Conv1D models are designed to capture temporal patterns in one-dimensional sequential data. In the case of climate data, where features such as temperature, pressure, humidity, etc., are recorded over time, a Conv1D architecture can effectively learn and extract patterns in the temporal sequence. Conv1D models use filters to scan local regions of the input sequence, allowing them to capture localized features. This is advantageous in climate data where certain local patterns or variations may be indicative of impending heatwaves. The ability to extract such localized features can enhance the model's sensitivity to critical patterns.

Convolutional layers in Conv1D models use parameter sharing, which helps in learning translation-invariant features. This is particularly useful in climate data, where the exact timing of temperature spikes or other features may vary. The model can learn to detect relevant features regardless of their specific temporal position.

Conv1D models have been successfully applied to various time-series data, including financial data and sensor readings. Their ability to capture temporal dependencies makes them a natural choice for climate-related time-series forecasting tasks, such as predicting heatwaves.

Conv1D architectures are computationally efficient compared to more complex models like LSTMs and GRU. This efficiency can be beneficial when working with large climate datasets, enabling faster training and evaluation.

Attention Mechanism

Long-term dependencies in time-series data can be captured by the model because of attention mechanisms. Attention mechanisms provide the ability to focus on relevant information at any time step, regardless of its temporal distance from the current step. Time-series forecasting models' interpretability is improved by attention mechanisms. By visualizing the attention weights, users can understand which parts of the historical data are most influential in making predictions. This interpretability helps in gaining insights into the driving factors behind the forecasts, making it easier to validate and trust the model's predictions. Attention mechanisms have revolutionized time-series forecasting by allowing models to effectively capture dependencies, handle irregular patterns, and provide interpretable forecasts. By assigning varying weights to different elements of the input sequence, attention mechanisms enable models to focus on relevant information and make accurate predictions [32].

3.7 Heat-Index

The heat index, also known as the apparent temperature, is a measure of how hot the human body feels when relative humidity is combined with the air temperature. It considers the effect of humidity on the body's ability to cool itself through perspiration which is very significant to measure how hot it feels. When the air temperature is high and relative humidity is also high, perspiration evaporates more slowly and makes it harder for the body to cool down. As a result, the heat index increases, and the body feels warmer than the actual air temperature. On the other hand, when relative humidity is low, perspiration evaporates more quickly, allowing the body to cool down, and the heat index decreases.

The formula below approximates the heat index in degrees Celsius, to within ± (0.7 °C). It is the result of a multivariate fit (temperature equal to or greater than 27 °C and relative humidity equal to or greater than 40%) to a model of the human body [33].

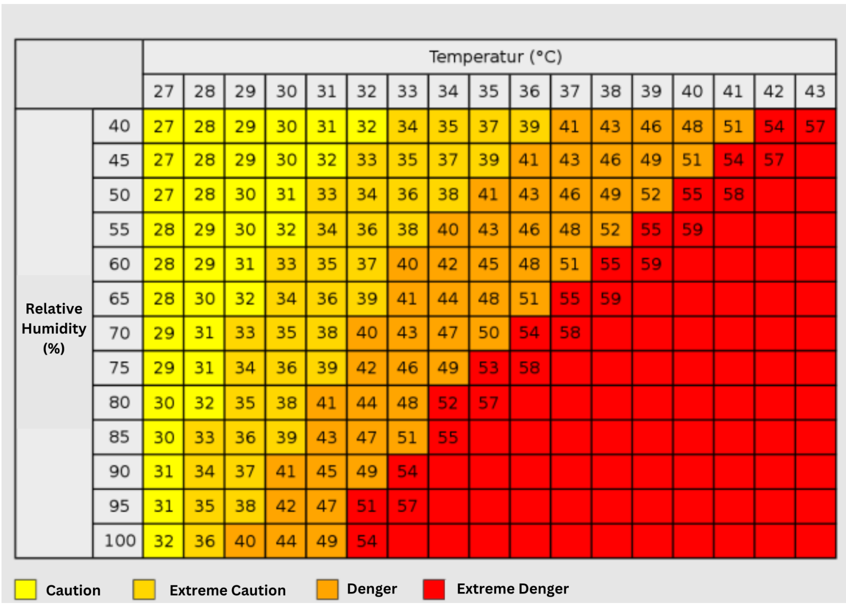
$$HI = c_1 + c_2T + c_3R + c_4TR + c_5T^2 + c_6R^2 + c_7T^2R + c_8TR^2 + c_9T^2R^2 \tag{3.15}$$

The following coefficients can be used to determine the heat index when the temperature is given in degrees Celsius, where

HI = heat index (in degrees Celsius)
T = ambient dry-bulb temperature (in degrees Celsius)
R = relative humidity (percentage value between 0 and 100)
$c_1 = -8.78469475556$
$c_2 = 1.61139411$
$c_3 = 2.33854883889$
$c_4 = -0.14611605$
$c_5 = -0.012308094$
$c_6 = -0.0164248277778$
$c_7 = 2.211732 \times 10^{-3}$
$c_8 = 7.2546 \times 10^{-4}$
$c_9 = -3.582 \times 10^{-6}$

The heat index is calculated based on the air temperature and relative humidity. A heat index chart or calculator is often used to determine the heat index value. For example, if the air temperature is 37.8°C and the relative humidity is 55%, the heat index would be 51.1°C. It's important to note that exposure to direct sunlight can increase the heat index by up to 8.3°C. The heat index values are classified into different categories based on their potential impact on the body:

- Caution (26.7°C - 32.2°C): Fatigue is possible with prolonged exposure or physical activity.
- Extreme Caution (32.2°C - 39.4°C): Heatstroke, heat cramps, or heat exhaustion are possible with prolonged exposure or physical activity.
- Danger (39.4°C - 51.1°C): Heat cramps or heat exhaustion are likely, and heatstroke is possible with prolonged exposure or physical activity.
- Extreme Danger (51.1°C or higher): Heatstroke is highly likely.



Relative Humidity (%)	Temperatur (°C)																
	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43
40	27	28	29	30	31	32	34	35	37	39	41	43	46	48	51	54	57
45	27	28	29	30	32	33	35	37	39	41	43	46	49	51	54	57	
50	27	28	30	31	33	34	36	38	41	43	46	49	52	55	58		
55	28	29	30	32	34	36	38	40	43	46	48	52	55	59			
60	28	29	31	33	35	37	40	42	45	48	51	55	59				
65	28	30	32	34	36	39	41	44	48	51	55	59					
70	29	31	33	35	38	40	43	47	50	54	58						
75	29	31	34	36	39	42	46	49	53	58							
80	30	32	35	38	41	44	48	52	57								
85	30	33	36	39	43	47	51	55									
90	31	34	37	41	45	49	54										
95	31	35	38	42	47	51	57										
100	32	36	40	44	49	54											

Figure 3.6: Heat Index [7]

These classifications help to understand the risks associated with different heat index values and take appropriate precautions to stay safe during hot and humid conditions [34]. The relation between heatwaves and the heat index is significant, as both concepts are used in assessing and understanding the impact of high temperatures on human health and well-being. The heat index quantifies how hot it feels to the human body by combining air temperature and humidity. Similarly, heat waves denote periods of exceptionally high temperatures. Both concepts focus on assessing the perceived temperature experienced by people during periods of elevated heat. Higher Heat Index values correspond to higher discomfort levels and indicate more intense heat wave conditions [8].

Chapter 4

Methodology

4.1 Proposed Methodology

The proposed methodology for developing deep learning models to predict maximum temperature and humidity using weather data from the BMD. The preparation of the dataset and its handling are entirely carried out in a scientific Python development environment (Colab) notebook. The model architectures are of a sequential type as shown in fig. 4.1, and it is developed using a Keras library, built on the top of tensorfow used for large-scale DL algorithms.

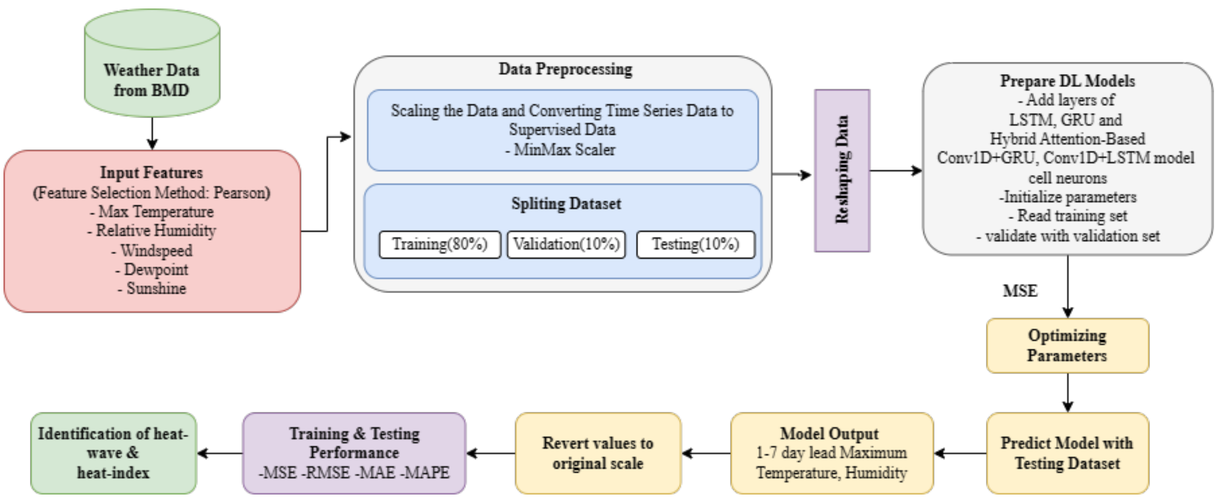


Figure 4.1: Methodology Architecture

The raw weather data undergoes scaling and conversion to a supervised format suitable for time series analysis using the MinMax Scaler technique. Additionally, the dataset is strategically split into training (80%), validation (10%), and testing (10%) subsets to facilitate effective model training and evaluation. These models encompass a diverse range of architectures, including Long Short-Term Memory (LSTM) networks, GRU and hybrid attention-based Conv1D+GRU and Conv1D+LSTM model cells are employed to capture complex patterns in the data. The deep learning models are meticulously trained by initializing parameters, feeding the training set, and continuously validating against the validation set. This iterative process ensures the models effectively learn the underlying patterns and relationships within the data.

To enhance model performance, fine-tune the model's hyperparameters based on the validation set's performance, enabling optimal configurations for accurate forecasting. Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE), are employed to assess the models' accuracy and robustness for predicting heat-wave and heat-index variables. To facilitate interpretability and practical application, the predicted values are reverted to their original scales, ensuring the forecasts align with real-world measurements and observations. As a final step, the models are tested on the reserved testing dataset to assess their performance on new, unseen data.

4.2 Data Collection

The dataset used in this research is a private dataset that has been collected from the Bangladesh Meteorological Department(BMD). We have used the historical observed weather data of Bangladesh from 1996 to 2020. The data has been previously used in several studies related to weather.

4.3 Data Description

We have selected district-wise data including Dhaka, Rajshahi, Bogra, Dinajpur, Jessore and Khulna. We have maintained one CSV file for the dataset. In this dataset, we have selected 3 months- March, April, and May data as in Bangladesh it is considered to be a hot season. Also, we have learned from the studies that the previous heatwave occurrences happened in these months. From the fig. 4.2, March, April, and May represent the pre-monsoon season in Bangladesh. During this period, the country experiences rising temperatures and increased humidity as it transitions from winter to the monsoon season. The pre-monsoon season is characterized by a gradual increase in temperature, making it a critical time to assess the likelihood of heatwaves. March, April, and May are characterized by a significant increase in temperatures in Bangladesh. These months mark the onset of warmer weather, contributing to the potential for extreme heat events. Historical meteorological data indicates that heatwaves in Bangladesh often occur during the pre-monsoon season. Analyzing this specific period provides insights into the conditions leading to heatwave events. The selection of March, April, and May for predicting heatwaves in Bangladesh is grounded in the climatic dynamics of the region, the historical occurrence of extreme heat events during this period, and the practical implications for agriculture, public health, and disaster management. Analyzing these months provides a focused and relevant assessment of the conditions leading to potential heatwaves, aiding in effective decision-making and mitigation strategies.

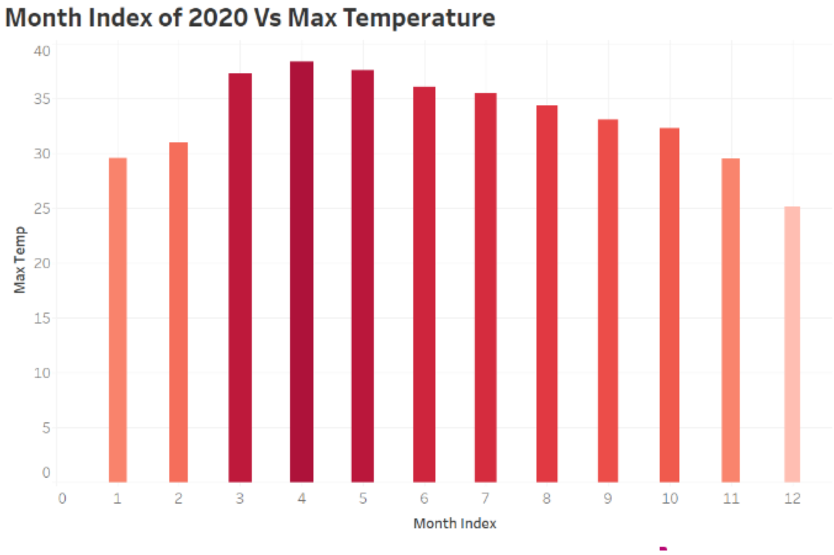


Figure 4.2: Exhibit 1 month-wise max temperature

4.4 Feature Selection

In our dataset we have features- maximum temperature, pressure, humidity, dew point temperature, sunshine, wind speed, rainfall. After applying the Pearson correlation method we have observed in fig. 4.3 that pressure and rainfall are weakly correlated to maximum temperature. Humidity is also weakly correlated to maximum temperature but it has some influence to predict heat index. Other features are positively related to maximum temperature. So the selected features are described below.

Max Temp (°C)
Maximum temperature is a key indicator of heatwaves. High temperatures are a primary characteristic of heatwave events. By including maximum temperature as a feature, the model can learn to associate spikes or sustained periods of high temperatures with the occurrence of heatwaves.

Humidity (%)
Humidity is an important factor in determining the discomfort associated with high temperatures. High humidity levels can exacerbate the impact of heat waves. Including humidity

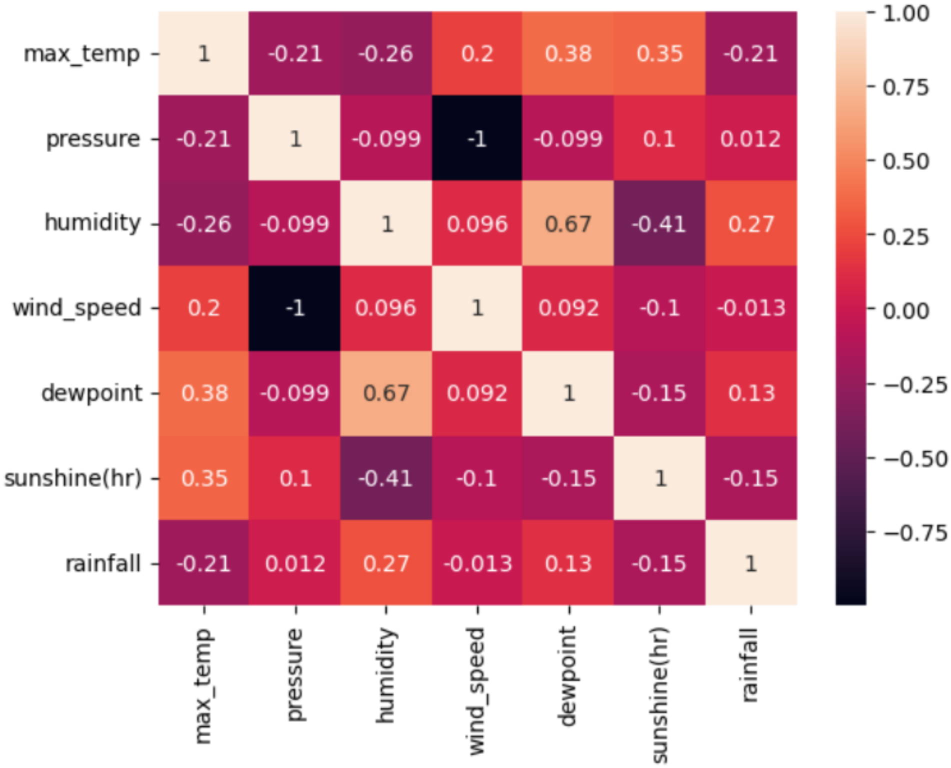


Figure 4.3: Correlation matrix between features

as a feature allows the model to consider the combined effect of temperature and humidity, providing a more comprehensive understanding of heatwave conditions.

Windspeed (m/s)

Wind speed refers to the rate at which air is moving horizontally past a specific point on the Earth's surface. Higher wind speeds can improve ventilation by distributing heat and reducing the buildup of hot air in a specific location. If the wind brings warmer air down from higher altitudes, it may increase the heat at the surface. On the other hand, mixing colder air downward can help reduce the impacts of a heatwave.

Dew point temperature(°C)

Dew is formed when the relative humidity hits 100%. The temperature at which water molecules saturate the air is referred to as the dew point. Warmer air can store more water molecules, and when it cools, that warm air loses water vapour through condensation. A higher dew point signifies more moisture in the air, resulting in oppressively humid conditions with the possibility for cloud and precipitation.

Sunshine (hr)

Sunshine duration is crucial for heatwave prediction as it directly affects the amount of solar radiation reaching the Earth's surface. Longer durations of sunshine contribute to higher temperatures. By incorporating sunshine hours as a feature, the model can capture the impact of solar radiation on temperature patterns, aiding in the identification of potential heatwave conditions.

4.5 Data Pre-processing

The first step is to prepare the weather dataset for the deep learning models. This involves framing the dataset as a supervised learning problem and normalizing the input variables. The formulation is to Predict the heatwave which is based on maximum temperature for the next 7 days based on the weather conditions. Identify the days that meet the heatwave criteria (e.g., three consecutive days with one or two variables exceeding the 95th percentile) [35]. The values obtained from CSV files are normalized between 0 and 1 to avoid the problem of scaling between different input features. Normalization typically refers to transforming the data to have a mean of 0 and a standard deviation of 1, while scaling generally means bringing the values into a specific range. Normalization, often achieved with tools like MinMaxScaler or StandardScaler, is essential to bring data to a common scale. In our context, we've utilized MinMaxScaler to transform the data into a range of [0, 1]. This is particularly crucial in neural networks to enhance convergence and prevent any single feature from dominating the learning process [36].

Chapter 5

Experimental Setup and Result Analysis

5.1 Experimental Setup

5.1.1 Train, Validation and Test sets

The number of days assigned for training the model is 80%, 10% for validation and 10% for testing for each city from the dataset. The first 1840 days are assigned for the training set and the next 460 days for validation and testing.

5.1.2 Windowing the data

The train and validation set data were split into a window of the previous seven days, with the following seven days serving as the target variable for the model being trained on, to predict individual features using separate models. A 2-D array of 7 days weather data are features of the dataset and an array of corresponding 7 days max-temperature and humidity data is the target variable to be trained on separate models. The 7 days maximum temperature and humidity are predicted by the separated model on the test set.

Suppose we want to predict the future 7 days' maximum temperature, for this the model will see the data from the previous 7 days. Considering today's features are t day, then the features of the previous 7 days will be : (t-1) day, (t-2) day, (t-3) day, (t-4) day, (t-5) day, (t-6) day, (t-7) day. Similarly the predicted max-temperature for next 7 days will be: t day, (t+1) day, (t+2) day, (t+3) day, (t+4) day, (t+5) day, (t+6) day. The model will learn the data from (t-7) day to (t-1) day and predict the temperature of t day, then it will add the predicted t day to the previous data and again will learn the sequence from (t-6) day to t day for predicting (t+1) day. The window will be sliding like this and in t day we will have the future 7 days maximum temperature.

The train and validation set data were split into a 7-day window for the purpose of predicting multitarget characteristics. The model was trained on a sliding window of 7 days of data in order to predict the subsequent 8th day. A 2-D array of 7 days weather data are features of the dataset and an array of corresponding 8th day's max-temperature and humidity data is the target variable to be trained on. The 8th day's maximum temperature and humidity are predicted by the model on the test set.

5.1.3 LSTM Model

To prepare the model for a single feature, two layers are added to form a Sequential LSTM Neural Network model using the Keras library. The first layer consists of 50 neurons and the second layer consists of 32 neurons. Dropout regularization with a rate of 0.3 is applied after the second LSTM layer. The third layer is a dense layer of 16 neurons with the ReLU activation function. A fourth layer of a single neuron is added for the output from the neural network. For multitarget feature prediction, the last layer is a dense layer of two neurons for predicting two features: Maximum Temperature and Humidity.

5.1.4 GRU model

To prepare the model for a single feature, two layers are added to form a Sequential GRU Neural Network model using the Keras library. The first layer consists of 64 neurons and the second layer consists of 32 neurons. Dropout regularization with a rate of 0.3 is applied after the second GRU layer. The third layer is a dense layer of 16 neurons with the ReLU activation function. A fourth layer of a single neuron is added for the output from the neural network.

5.1.5 Attention-based Conv1D+GRU and Conv1D+LSTM

The proposed model architecture consists of two 1D convolutional layers with 32 filters and 3 kernel size with ReLU activation and batch normalization. An attention mechanism is integrated into the model using an attention block. The attention mechanism is applied after the initial feature extraction, comprising Conv1D layers for feature transformation and before the GRU layers allowing the model to learn to attend to relevant features while processing the sequential data and capturing temporal dependencies in the data. Two GRU layers are stacked(first layer with 50 neurons and second layer with 32 neurons), with the first returning sequences to maintain temporal context. Dropout regularization with a dropout rate of 0.3 is applied to prevent overfitting by randomly dropping out a fraction of the units during training. The final output layer consists of a single neuron, with a kernel regularizer applied to mitigate overfitting. The Conv1D+LSTM model has the same layers too.

5.1.6 Optimization

Adam optimizer is used for optimization. Mean Squared Error (MSE) loss function is used. EarlyStopping callback is implemented to monitor validation loss. Training will stop if the validation loss does not improve for 20 consecutive epochs. The models are trained using the training data for 100 epochs with a batch size of 64.

5.1.7 Evaluation Metrics

Evaluation metrics are quantitative measures used to assess the performance and effectiveness of a statistical or machine-learning model. These metrics provide insights into how well the model is performing and help in comparing different models or algorithms [37].

Mean Squared Error(MSE)

The MSE is a measure of the quality of an estimator. As it is derived from the square of Euclidean distance, it is always a positive value that decreases as the error approaches zero [38].

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{5.1}$$

Root Mean Squared Error (RMSE)

(RMSE) is one of the two main performance indicators for a regression model. It measures the average difference between values predicted by a model and the actual values. It provides an estimation of how well the model can predict the target value (accuracy) [39].

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{5.2}$$

Mean Absolute Error(MAE)

Mean absolute error (MAE) is measured as the average absolute difference between the predicted values and the actual values and is used to assess the effectiveness of a regression model [40].

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - x_i| \tag{5.3}$$

where y_i is the prediction and x_i is the true value.

Mean Absolute Percentage Error(MAPE)

The mean absolute percentage error (MAPE), also known as mean absolute percentage deviation (MAPD), is a measure of the prediction accuracy of a forecasting method in statistics. It usually expresses the accuracy as a ratio defined by the formula:

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{A_i - F_i}{A_i}\right| \times 100\% \tag{5.4}$$

where A_t is the actual value and F_t is the forecast value. Their difference is divided by the actual value A_t. The absolute value of this ratio is summed for every forecasted point in time and divided by the number of fitted points n [41].

5.2 Result Analysis

5.2.1 Model Performance

The results from the multitarget prediction and single feature prediction are compared in Tables 5.1 and 5.2, respectively. The hybrid model achieved the highest MSE of 55.51 in the multitarget prediction scenario, whereas the LSTM model achieved MSE of 39.42. The MSE of 28.11 for the GRU model indicates comparatively greater performance. Comparing the LSTM and hybrid models to the GRU model (5.30), the RMSE values of 6.28 and 7.45, respectively, were higher. These increased error values show that the models had trouble concurrently predicting several target variables with accuracy.

Table 5.1: Result of Multitarget prediction

Multitarget prediction	LSTM	GRU	Hybrid
MSE	39.42	28.11	55.51
RMSE	6.28	5.30	7.45

Table 5.2: Result of Single Feature prediction

Single Feature prediction	LSTM	GRU	Hybrid
MSE	6.33	5.51	4.32
RMSE	2.52	2.35	2.08

On the other hand, in the single feature prediction, the error values across all models were notably lower. The MSE values for the LSTM, GRU, and hybrid models were 6.33, 5.51, and 4.32, respectively. Likewise, the RMSE values were also lower for single-feature prediction compared to multitarget prediction, with the LSTM, GRU, and hybrid models showing values of 2.52, 2.35, and 2.08, respectively.

These comparisons suggest that the models' performance was better in the single feature prediction task than in the multitarget prediction task. The multitarget prediction scenario's higher error values imply that the models had trouble correctly predicting many target variables at once, which got less ideal results. Therefore, more research or model development for multitarget prediction might be necessary to enhance performance.

5.2.2 Comparison of Models

The table 5.3 presents a comparative analysis of four different models based on their performance metrics: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). It shows the evaluation for the prediction of seven days lead maximum temperature.

Table 5.3: Comparative Analysis of the models(Dhaka)

Models	Metrics	1d	2d	3d	4d	5d	6d	7d
Attention-based Conv1D+GRU	RMSE	2.08	2.48	2.60	2.76	2.78	2.58	2.71
	MAE	1.40	1.77	1.91	2.10	2.24	2.01	2.17
	MAPE	4.53%	5.69%	6.07%	6.59%	6.87%	6.20%	6.65%
Attention-based Conv1D+LSTM	RMSE	2.23	2.76	2.82	3.26	3.37	2.81	2.90
	MAE	1.69	2.01	2.00	2.50	2.56	2.14	2.37
	MAPE	5.32%	6.46%	6.43%	7.59%	7.75%	6.63%	7.16%
GRU	RMSE	2.35	2.11	2.26	2.30	2.37	2.30	2.36
	MAE	2.11	1.66	1.73	1.68	1.74	1.72	1.79
	MAPE	6.40%	5.22%	5.46%	6.34%	6.49%	6.88%	6.71%
LSTM	RMSE	2.52	2.66	2.47	2.54	2.93	2.73	2.74
	MAE	2.23	2.30	2.02	2.03	2.33	2.23	2.25
	MAPE	6.65%	6.88%	6.15%	6.24%	7.05%	6.82%	6.83%

Attention-based Conv1D+GRU model shows relatively low RMSE values across all lead times, ranging from 2.08 to 2.78. The MAPE values for this model range from 4.53% to 6.87%, indicating relatively low error rates. Overall, it performs better compared to the other models. While the attention-based Conv1D+LSTM model performs reasonably well, it generally shows higher RMSE, MAE, and MAPE compared to the Attention-based Conv1D+GRU. The RMSE values range from 2.23 to 3.37, which are higher than those of the Conv1D+GRU model. The GRU model's performance falls between the Conv1D+LSTM and Conv1D+GRU models, with slightly higher error metrics. Among the models evaluated, LSTM generally performs the poorest, with higher error metrics across all lead times.

5.2.3 Results for each district

Temperature prediction is essential for numerous sectors, including agriculture, energy management, and disaster preparedness. Evaluate the performance of temperature prediction models for six districts: Dhaka, Rajshahi, Bogra, Khulna, Dinajpur and Jessor. The evaluation metrics include Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) at various lead times. Factors such as geographic location, local climate patterns, and data quality may influence the performance of prediction models. Districts with more stable weather patterns or better data availability may exhibit lower prediction errors. Prediction errors tend to increase with longer lead times. This trend is consistent across all districts, indicating the inherent difficulty in accurately forecasting temperature over extended time horizons. The performance of temperature prediction models varies across districts and lead times. Generally, districts like Dhaka and Khulna demonstrate relatively lower prediction errors compared to districts like Rajshahi, Bogra, Dinajpur, and Jessore. These findings underscore the importance of tailoring prediction models to specific geographical locations and continuously refining them to enhance accuracy, particularly for longer-term forecasts.

The results of Attention-based Conv1D + GRU models for all districts are given in table 5.4.

Table 5.4: Results of all districts(Attention-based Conv1D+GRU)

District	**Metrics**	**1d**	**2d**	**3d**	**4d**	**5d**	**6d**	**7d**
Dhaka	RMSE	2.08	2.48	2.60	2.76	2.78	2.58	2.71
	MAPE	4.53%	5.69%	6.07%	6.59%	6.87%	6.20%	6.65%
Rajshahi	RMSE	2.28	3.04	3.19	3.12	3.26	3.33	3.08
	MAPE	5.26%	7.18%	7.66%	7.51%	8.01%	7.86%	7.50%
Bogra	RMSE	2.64	3.08	3.55	3.57	3.11	3.55	3.68
	MAPE	6.38%	7.70%	8.69%	9.25%	8.08%	9.22%	9.64%
Khulna	RMSE	2.00	2.40	2.41	2.76	2.63	2.70	2.84
	MAPE	4.30%	5.26%	5.36%	6.15%	6.05%	6.32%	6.61%
Dinajpur	RMSE	2.60	2.99	3.00	3.05	2.98	3.29	3.19
	MAPE	6.38%	7.60%	7.97%	8.18%	7.82%	8.29%	8.40%
Jessore	RMSE	1.97	2.35	2.70	3.02	2.93	3.11	2.66
	MAPE	4.26%	5.39%	5.94%	6.73%	6.86%	7.25%	6.08%

The results of Attention-based Conv1D + LSTM models for all districts are given in table 5.5.

Table 5.5: Results of all districts(Attention-based Conv1D+LSTM)

District	Metrics	1d	2d	3d	4d	5d	6d	7d
Dhaka	RMSE	2.23	2.76	2.82	3.26	3.37	2.81	2.90
	MAPE	5.32%	6.46%	6.43%	7.59%	7.75%	6.63%	7.16%
Rajshahi	RMSE	2.55	3.06	3.45	3.76	3.59	3.66	3.81
	MAPE	6.05%	7.40%	8.59%	9.00%	8.99%	9.13%	9.57%
Bogra	RMSE	2.71	2.97	3.66	3.63	4.40	3.76	3.51
	MAPE	6.48%	7.53%	9.48%	9.54%	11.47%	9.78%	9.16%
Khulna	RMSE	2.03	2.32	2.52	2.80	2.90	3.08	2.90
	MAPE	4.44%	5.11%	5.54%	6.23%	6.77%	7.00%	6.81%
Dinajpur	RMSE	2.63	3.25	3.22	3.47	3.78	3.61	3.66
	MAPE	6.62%	8.37%	8.45%	9.09%	9.83%	9.54%	9.80%
Jessore	RMSE	2.14	2.62	3.32	2.85	2.69	2.84	2.83
	MAPE	5.02%	5.99%	7.11%	6.50%	6.07%	6.40%	6.58%

The results of GRU models for all districts are given in table 5.6.

Table 5.6: Results of all districts(GRU)

District	Metrics	1d	2d	3d	4d	5d	6d	7d
Dhaka	RMSE	2.35	2.41	2.76	2.80	2.75	2.90	2.96
	MAPE	6.40%	5.62%	6.58%	6.64%	6.51%	6.78%	6.81%
Rajshahi	RMSE	2.33	2.95	3.03	3.77	3.50	3.76	3.58
	MAPE	5.32%	6.88%	7.38%	9.57%	8.58%	9.17%	8.80%
Bogra	RMSE	2.29	2.95	2.83	3.23	3.12	3.68	3.25
	MAPE	6.54%	7.32%	7.37%	8.44%	8.12%	9.37%	8.43%
Khulna	RMSE	1.85	2.36	2.53	2.81	2.91	2.76	2.97
	MAPE	4.41%	5.55%	5.77%	6.48%	7.07%	6.54%	6.73%
Dinajpur	RMSE	2.48	2.92	3.00	3.62	3.05	3.12	3.04
	MAPE	6.19%	7.17%	8.01%	9.54%	7.95%	8.12%	7.85%
Jessore	RMSE	2.26	2.89	2.53	3.11	2.81	2.84	2.66
	MAPE	5.69%	7.17%	5.67%	7.22%	6.34%	6.54%	6.28%

The results of LSTM models for all districts are given in table 5.7.

Table 5.7: Results of all districts(LSTM)

District	Metrics	1d	2d	3d	4d	5d	6d	7d
Dhaka	RMSE	2.52	2.66	2.47	2.54	2.93	2.73	2.74
	MAPE	6.65%	6.88%	6.15%	6.24%	7.05%	6.82%	6.83%
Rajshahi	RMSE	2.50	2.74	2.92	2.93	4.31	3.58	3.31
	MAPE	6.06%	6.67%	7.04%	7.08%	9.58%	8.33%	7.82%
Bogra	RMSE	2.26	2.94	2.65	3.21	2.68	2.80	2.97
	MAPE	6.00%	7.47%	6.84%	8.43%	7.12%	7.37%	7.75%
Khulna	RMSE	1.94	2.30	2.57	2.57	2.85	2.59	2.47
	MAPE	4.50%	5.44%	6.30%	5.90%	7.06%	5.92%	5.89%
Dinajpur	RMSE	2.37	2.60	2.79	3.00	3.19	3.02	3.14
	MAPE	5.87%	6.45%	7.39%	7.76%	8.41%	7.85%	8.15%
Jessore	RMSE	2.19	2.88	2.59	2.98	2.95	2.76	2.75
	MAPE	5.40%	7.13%	6.08%	7.18%	7.01%	6.38%	6.42%

Observing the result, a gradual decrease in performance metrics as the lead time increases for most districts. This trend suggests that predicting maximum temperatures becomes increasingly challenging as we look further ahead into the future. For Dhaka, Rajshahi, Khulna and Jessore Attention-based Conv1D + GRU model shows less RMSE and MAPE for consecutively days. On the other hand, for Bogura GRU shows less RMSE and MAPE for consecutively days. Also for Dinajpur LSTM shows better results from the other models. But overall comparing the performance Attention-based Conv1D + GRU gives better performance in the cases of all districts if consecutive days are considered.

5.2.4 Comparison between proposed DL-based model and WRF

With the increase in advancement in technology since the twentieth century, access to weather forecasts has been drastically increased. During the period from May 10th to May 20th, 2019, the predicted temperature of our hybrid model and WRF model has been shown in table 5.8. Analyzing temperature predictions generated by the Weather Research and Forecasting (WRF) model and hybrid model from 2018 to 2019 have been compared. WRF model gives 2.36 MAE and 3.00 RMSE while our hybrid model gives 1.42 MAE and 1.66 RMSE for Dhaka district which is less than the WRF model. From the comparison, it's evident that the WRF model consistently predicts higher maximum temperatures compared to the observed data. The hybrid model(attention-based Conv1D+GRU), on the other hand, falls between the WRF predictions and the observed data, often closer to the observed values.

Table 5.8: Comparison of WRF and Attention-based Conv1D + GRU

Date	WRF(°C)	Observed(°C)	Attention-based Conv1D+GRU(°C)
2019-05-10	38.36	36.2	34.32
2019-05-11	40.52	35.6	33.89
2019-05-12	36.50	35.8	35.31
2019-05-13	39.19	35.5	35.08
2019-05-14	35.92	34.4	34.70
2019-05-15	37.58	35.2	34.53
2019-05-16	35.82	35.4	33.44

This discrepancy between the WRF model and observed data could be due to several factors, including model biases, inaccuracies in input data, or limitations in the model's ability to accurately simulate real-world conditions. While the WRF model may sometimes overestimate temperatures, the hybrid model offers a more realistic estimate, aligning better results in temperature predictions that are closer to what actually observed in the real world.

5.2.5 Heatwave Detection

Detecting heat waves involves a special approach where we analyze temperature trends to understand potentially hazardous weather conditions. To detect potential heat waves, meteorologists analyze the temperature patterns over consecutive days. They generally take 36°C as the threshold value for Bangladesh's maximum temperature. To set a standard for high temperatures, we utilize historical weather data spanning from 1996 to 2020. Through this analysis, it is determined that the 95th percentile of observed temperatures is 36°C, which matches with the observation of BMD meteorologists. So it has been used as a threshold for detecting heat waves from the predicted temperatures, indicating when temperatures exceed the normal range and enter into heatwave territory.

Employing advanced deep learning techniques to forecast such as GRU, LSTM, Conv1D combined with GRU, and Conv1D combined with LSTM with an attention mechanism we predict temperatures for the upcoming 7 days. Once the forecasts are generated, the focus turns to identifying potential heatwave conditions. The patterns where the temperature remains consistently high for extended periods. Specifically, if there are 3-4 consecutive days from the predicted 7 days where temperatures equal or exceed the threshold of 36°C, it triggers the detection of a potential heatwave. This approach of heatwave detection allows us to issue timely warnings and advisories, enabling communities to prepare and take necessary precautions.

5.2.6 Heat-Index

The Heat index is particularly important in assessing heat-related discomfort and health risks, especially during hot and humid weather conditions. Using the predicted values of maximum temperature and humidity obtained from our hybrid model by calculating the Heat Index. The Heat Index provides an indication of a hot day and how hot it would feel considering both temperature and humidity. This calculation incorporated the predicted temperature and humidity values, along with specific constants to determine the perceived temperature. The calculated Heat Index values were analyzed to assess potential heat-related discomfort and health risks. The results were compared with established thresholds and guidelines for heat risk assessment.

Table 5.9: Heat Index

Days	Temperature(°C)	Humidity(%)	Heat Index(°C)
1	32.25	78.35	44.56
2	31.05	73.37	38.78
3	30.74	75.93	38.70
4	29.87	76.79	36.41
5	30.55	84.31	40.84
6	29.37	80.99	36.04
7	32.90	77.52	46.55

When the prediction of the attention-based Conv1D + GRU is less than 36°C, it can not be considered as a heat wave day, but for higher humidity human body will feel a higher temperature which can be considered as a hot day. For example, in table 5.9 there has been shown a consecutive 7-day ahead prediction for 26th April 2018. If the temperature is 32.25°C and humidity is 78.35% then the human body will feel 44.56°C (heat index temperature). For this reason, it is needed to predict humidity alongside temperature.

From the discussion with the meteorologists of the Bangladesh Meteorological Department, there is still some research going on considering heat index temperature for predicting heat wave days. Yusha Anis et al. [42] proposed that the results of all heatwave events can be calculated based on the occurrence of the 97th percentile value against all recorded observations at one given point. The output is the 97th percent highest recorded Heat Index value, for that given point. But in Bangladesh, the 95th percentile is considered. So after predicting the heat index the 95th percentile value can be calculated and it can be considered as a threshold value to predict the probability of heat wave days.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

Being able to predict heatwave days accurately is crucial for getting ready for disasters, protecting people's health, and creating plans that work well with the changing climate. This study presents the potential for multi-step-ahead (1-day to 7-day) prediction of daily maximum temperature and humidity separately thereafter exploring its potential to foresee the upcoming heatwave events. The research included the evaluation of Long Short-Term Memory (LSTM), Gated Recurrent Unit(GRU), and Hybrid modes like attention-based Conv1D+LSTM and Conv1D+GRU. The Conv1D+GRU model gives the lowest RMSE(2.08), MAE(1.40) and MAPE(4.53%) which indicates a better performance for 1-day ahead prediction for Dhaka district. This model shows less MAPE for other districts too if consecutively days are considered. The accuracy of predictions given up to seven days in advance gradually decreases based on the districts' data variety. This allows for the prediction of heatwaves based on the observation of maximum temperature rise for three to four days(>36°C). The heat index that is calculated using the predicted temperature and humidity has the potential to predict future weather variations in Bangladesh.

6.2 Future Work

In the ongoing pursuit of advancing heatwave predictability in Bangladesh through deep learning methodologies, the findings of this study present opportunities for further exploration and enhancement within the realm of time series forecasting.

Exploring More Models

In the future, the Autoformer model can be used which can be effective for Time Series

Forecasting. Explainable AI(XAI) refers to a set of techniques and approaches designed to provide a clear and human-readable explanation for the decisions made by AI and machine learning models. Instead of giving an in-depth explanation of the model over the entire dataset, LIME concentrates on explaining the model's prediction for individual instances. LIME is a model-agnostic technique, meaning it can be applied to any machine learning model, including complex models like deep neural networks, which are often considered "black boxes." However, LIME provides explanations for individual instances and may not provide a comprehensive understanding of the model's global behavior. LIME can help determine the most important variables that contributed to a given heatwave prediction, such as highlighting that high temperatures, humidity are the key drivers for predicting a heatwave occurrence.

Dynamic Hyperparameter Optimization

Although preliminary optimization of hyperparameters has been undertaken, there exists room for a more comprehensive exploration into dynamic hyperparameter tuning techniques. This could involve adapting to the evolving nature of climate data and its influence on model performance within the context of time series forecasting. This could involve a more extensive search for optimal configurations to enhance model performance.

References

[1] "'alarming' heat wave threatens bangladesh's people and their food supply." https://news.mongabay.com/2023/04/alarming-heat-wave-threatens-bangladeshs-people-and-their-food-s
Accessed: 28 April 2023.

[2] O. Calzone, "An intuitive explanation of lstm." https://medium.com/@ottaviocalzone/an-intuitive-explanation-of-lstm-a035eb6ab42c, 2022.

[3] T. Muhammad, M. Irfan, A. Mustafa, A. Irtaza, Z. U. Rehman, S. Chaudhry, A. Abbas, S. A. Khan, and A. Malik, "Ggtr: An innovative framework for accurate and realistic human motion prediction," *arXiv preprint arXiv:2110.00215*, 2021.

[4] R. S. Srinivasamurthy, "Understanding 1d convolutional neural networks using multi-class time-varying signals," 2018.

[5] M. R. Keyvanpour and M. B. Shirzad, "Convolutional layer." https://www.sciencedirect.com/topics/computer-science/convolutional-layer, 2022.

[6] T. Nadeem, M. Irfan, A. Mustafa, A. Irtaza, Z. U. Rehman, S. Chaudhry, A. Abbas, S. A. Khan, and A. Malik, "Joint learning of temporal models to handle imbalanced data for human activity recognition," *IEEE Access*, 2020.

[7] "Heat index formula." https://www.toppr.com/guides/physics-formulas/heat-index-formula/.

[8] A. Awasthi, K. Vishwakarma, and K. C. Pattnayak, "Retrospection of heatwave and heat index," *Meteorology and Atmospheric Physics*, vol. 133, pp. 27–36, 2021.

[9] E. Robinson, https://www.overleaf.com/project/652be6ffbefff3c9719e444cShouro Dasgupta, and L. Letsch, "Adapting to the impacts of extreme heat in bangladesh's labour force." https://www.lse.ac.uk/granthaminstitute/news/

adapting-to-the-impacts-of-extreme-heat-in-bangladeshs-labour-f
2023.

[10] M. Billah, "How heatwaves are affecting livelihoods and food security." https://www.tbsnews.net/features/panorama/how-heatwaves-are-affecting-livelihoods-and-food-security-669834
2023.

[11] T. D. Star, "Labour productivity hit by extreme temperatures: Dhaka loses \$6b a year," *The Daily Star*, 2024. https://www.thedailystar.net/environment/climate-change/news/labour-productivity-hit-extreme-temperatures-dhaka-loses-6b-yea

[12] S. Karmakar and M. K. Das, "On the heat waves in bangladesh, their trends and associated large-scale tropospheric conditions," *Journal of Engineering Science*, vol. 11, no. 1, pp. 19–36, 2020.

[13] M. B. Rashid, S. S. Hossain, M. A. Mannan, K. M. Parding, H. O. Hygen, R. E. Benestad, and A. Mezghani, "Climate change projections of maximum temperature in the premonsoon season in bangladesh using statistical downscaling of global climate models," *Adv. Sci. Res.*, vol. 18, p. 99–114, 2021.

[14] S. Wang, W. Cai, Y. Tao, Q. Sun, P. Wong, X. Huang, and Y. Liu, "Unpacking the inter- and intra-urban differences of the association between health and exposure to heat and air quality in australia using global and local machine learning models," *Science of The Total Environment*, vol. 871, p. 162005, 2023.

[15] A. S. S.B.H.S. Asadollah, N. Khan *et al.*, "Prediction of heat waves using meteorological variables in diverse regions of iran with advanced machine learning models," *Stoch Environ Res Risk Assess*, vol. 36, p. 1959–1974, 2022.

[16] P. Li, Y. Yu, D. Huang, Z.-H. Wang, and A. Sharma, "Regional heatwave prediction using graph neural network and weather station data," *Geophysical Research Letters*, vol. 50, no. e2023GL103405, 2023.

[17] M. Park, D. Jung, S. Lee, and S. Park, "Heatwave damage prediction using random forest model in korea," *Applied Sciences*, vol. 10, no. 22, p. 8237, 2020.

[18] A. Kabbori, J. Antari, R. Iqdour, and Z. E. Abidine El Morjani, "Temperature prediction using time series time-delay neural networks," in *2019 7th International Renewable and Sustainable Energy Conference (IRSEC)*, pp. 1–4, 2019.

[19] N. Khan, S. Shahid, L. Juneng, K. Ahmed, T. Ismail, and N. Nawaz, "Prediction of heat waves in pakistan using quantile regression forests," *Atmospheric Research*, vol. 221, pp. 1–11, 2019.

[20] S. Juna, S. Narejo, and M. M. Jawaid, "Regional heatwave prediction using deep learning based recurrent neural network," in *2022 International Conference on Emerging Technologies in Electronics, Computing and Communication (ICETECC)*, pp. 1–5, 2022.

[21] A. Mane, N. Lekurwale, P. Maidamwar, P. Khobragade, and S. Dongre, "Artificial intelligence based heatwave intensity prediction model," in *2023 International Conference on IoT, Communication and Automation Technology (ICICAT)*, pp. 1–5, 2023.

[22] C. S. Bangera, P. S. Kotian, C. Dias, T. Divya, and G. Aithal, "Flood and heat wave prediction using weighted moving average, anomaly detection and k-nearest neighbours for the city of mangalore," in *2018 IEEE Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER)*, pp. 93–97, 2018.

[23] Z. Zhang and Y. Dong, "Temperature forecasting via convolutional recurrent neural networks based on time-series data," *2020*, p. 3536572.

[24] C. Olah, "Understanding lstm networks." `https://colah.github.io/posts/2015-08-Understanding-LSTMs/`, 2015.

[25] C. Olah, "Understanding convolutions," 2014.

[26] S. P. Katta, "Time series forecasting using conv1d-lstm: Multiple timesteps into future," 2021. Medium.

[27] S. K. Prasad and G. S. Koushik, "Time series forecasting using conv1d-lstm: Multiple timesteps into future," *IEEE*, vol. 19, pp. 408–412, 2021.

[28] Y. Li, B. Kong, W. Yu, and X. Zhu, "An attention-based cnn-lstm method for effluent wastewater quality prediction," *Applied Sciences*, vol. 13, no. 12, p. 7011, 2021.

[29] S. Wei and X. Bai, "An attention-based cnn-gru model for resident load short-term forecast," *IEEE Xplore*, 2021.

[30] National Center for Atmospheric Research (NCAR), "Weather Research and Forecasting (WRF) Model." `https://www.mmm.ucar.edu/models/wrf`, Accessed: 2024.

[31] J. Michalakes, J. Dudhia, D. Gill, T. Henderson, J. Klemp, and W. Skamarock, "The weather research and forecasting (wrf) model: Overview, system efforts, and future directions," *Bulletin of the American Meteorological Society*, vol. 98, no. 8, 2017.

[32] A. Vidhya, "Time series forecasting using attention mechanism," June 2023.

[33] "Heat index." https://en.wikipedia.org/wiki/Heat_index.

[34] "What is the heat index?." https://www.weather.gov/ama/heatindex.

[35] V. Mehandzhiyski, "How to pre-process time series data." https://365datascience.com/tutorials/time-series-analysis-tutorials/pre-process-time-series-data/, 2023.

[36] J. Brownlee, "How to scale data for long short-term memory networks in python." https://machinelearningmastery.com/how-to-scale-data-for-long-short-term-memory-networks-in-python 2019.

[37] T. Srivastava, "12 important model evaluation metrics for machine learning everyone should know." https://www.analyticsvidhya.com/blog/2019/08/11-important-model-evaluation-error-metrics/#:~:text=Evaluation%20metrics%20are%20quantitative%20measures,comparing%20different%20models%20or%20algorithms., 2023.

[38] "Mean squared error (mse)." https://www.probabilitycourse.com/chapter9/9_1_5_mean_squared_error_MSE.php.

[39] J. Frost, "Root mean square error (rmse)." https://statisticsbyjim.com/regression/root-mean-square-error-rmse/.

[40] Wikipedia contributors, "Mean absolute error." https://en.wikipedia.org/wiki/Mean_absolute_error, 2022. [Online; accessed 1-March-2022].

[41] Wikipedia contributors, "Mean absolute percentage error." https://en.wikipedia.org/wiki/Mean_absolute_percentage_error, 2022. [Online; accessed 1-March-2022].

[42] Y. Anis and A. Ali, "Developing heat index for heat wave hazard mapping: A case study of sindh province, pakistan," in *Seventh International Conference on Aerospace Science & Engineering (ICASE)*, IEEE, IEEE, 2021.

Appendix A

Codes

A.1 Getting Data for Each District

```
weather_data = weather_data[['date', 'station', 'max_temp', 'humidity',  'wind_speed','dewpoint', 'sunshine(hr)']]
dhaka = weather_data[weather_data['station'] == 'Dhaka'].dropna()
raj = weather_data[weather_data['station'] == 'Rajshahi'].dropna()
bogra= weather_data[weather_data['station'] == 'Bogra'].dropna()
khulna = weather_data[weather_data['station'] == 'Khulna'].dropna()
din = weather_data[weather_data['station'] == 'Dinajpur'].dropna()
jes = weather_data[weather_data['station'] == 'Jessore'].dropna()
```

Figure A.1: District wise data splitting

A.2 Data Supervision

```
def make_supervised(data_scaled, data):
    data_sup = series_to_supervised(data_scaled, n_in=7, n_out=7)
    drop_cols = ['var2(t)', 'var3(t)', 'var4(t)', 'var5(t)' ]

    for t in range(1, 7):
        for feat in range(1, 6):

            if not (feat == 1):
                drop_cols.append(f'var{feat}(t+{t})')
    data_sup = data_sup.drop(columns=drop_cols)

    return data_sup
```

Figure A.2: District wise data splitting

Generated using Undegraduate Thesis LaTeX Template, Version 1.4. Department of Computer Science and Engineering, Ahsanullah University of Science and Technology, Dhaka, Bangladesh.

This thesis was generated on Thursday 16^{th} May, 2024 at 6:08pm.