Eurasian J. Sci. Eng., 2024, 10(2), 16-33
https://doi.org/10.23918/eajse.v10i2p02

Published by Tishk International University
Available at: https://eajse.tiu.edu.iq/submit/

RESEARCH ARTICLE Open Access

Forecasting Weather Using Deep Learning from the Meteorological Stations Data: A Study of Different Meteorological Stations in Kaski District, Nepal

Supath Dhital [1,2] *, Kapil Lamsal [2], Sulav Shrestha [2], and Umesh Bhurtyal [2]

[1] Department of Geography and Environment Studies, The University of Alabama, Tuscaloosa AL-USA

[2] Department of Geomatics Engineering, Faculty of Engineering, Tribhuvan University, Pokhara-Nepal

Article History
Received: 10.07.2023
Revised: 30.03.2024
Accepted: 04.04.2024
Published: 28.04.2024
Communicated by: Asst. Prof. Dr. Abubakar H. Ashir
**Email address:*
sdhital@crimson.ua.edu
**Corresponding Author:*

Abstract:
Contemporarily, one of the most pressing concerns is reliable and rapid weather forecasting. In Nepal, the Department of Hydrology and Meteorological uses a numerical modeling approach to forecast the weather, which is tardy and requires high-end equipment to process the information, so a deep learning approach will be the best alternative. This project aims to forecast the next 2-hour Precipitation and Air Temperature for Pokhara Domestic Airport meteorological station and the next day's Precipitation, Maximum and Minimum Air Temperature forecast for Lumle, Begnas, and Lamachaur meteorological station, total of four meteorological stations of the Kaski District, Nepal using Long Short-Term Memory (LSTM): a Recurrent Neural Network (RNN) and deploy the outputs through the web portal. The four hourly parameters: Rainfall, Relative Humidity (R.H), Wind Speed, and Air Temperature, were used for modeling the airport station forecast, whereas Rainfall, Relative Humidity (R.H), Maximum and Minimum Temperature were used for modeling the Begnas and Lumle station forecast and only Precipitation data was used for Lamachaur station. Averaging and linear interpolation techniques were used to fill out the missing values and outliers were detected using Box Plot and replaced with threshold value for each parameter. Stochastic Gradient Descent and Adam optimizer are used to optimize the LSTM model. Among all the models prepared, Root Mean Square Error (RMSE) values range from 0.58 to 4.08 for the precipitation model and from 0.16 to 0.82 for the air temperature model, and Mean Absolute Error (MAE) values range from 0.21 to 2.87 for the precipitation model and from 0.12 to 0.64 for air temperature model were the values of the final model that indicates better accuracy for air temperature. The R^2 values range from 0.89 to 0.99, indicating the train and test data were fitted to the model really well.

Keywords: *Weather Forecast; Deep Learning; Long Short-Term Memory (LSTM); Meteorological Data; Precipitation; Air Temperature*

1. Introduction

Weather is defined as a persistent, multidimensional, dynamic, and data-intensive process that is characterized by variables such as temperature, humidity, precipitation, wind, and cloud cover at a specific time and place [1], which shows the atmospheric status of the Earth at different times and places. Knowing the weather extremities such as cyclones, thunderstorms, flooding, and heavy rains [2] in the past will help to avoid and mitigate them with less loss. In the context of Nepal, the 72- hour based short-range weather forecasting system was initiated by using the Numeric Weather Prediction (NWP) system and has been delivering a periodical Climate Bulletin to the public through its website (https://www.dhm.gov.np/bulletins) [3]. In Nepal, observed weather parameter was provided by 6 aero-synoptic, 9 synoptic, 20 sediments, 22 agro-meteorological, 68 climatic, 154 hydro-metric, and 337 precipitation stations [4].

Due to the diverse changes in geological terrains, rapid urbanization, and climate change, the prediction of precipitation is getting more complex and has a high chance of containing ambiguity [5]. Precipitation prediction plays a vital role in the simulation of hydrological activity, so predicting the precipitation to analyze several geomorphological activities [6] is also a vital application. Melting the glaciers in the Himalayas, probabilities of extreme weather conditions, and several natural disasters may occur due to the rising temperature [7], which is so devastating. Air temperature plays a crucial impact in measuring the greenhouse effect, solar radiation estimations, air pollution [8], [9], and so many other effects, so knowing it primarily helps to mitigate the various problems. Air temperature and precipitation, basically rapid weather forecasting, is a very crucial climatic factor required for many different applications in domains like energy, industry, environment, tourism, agriculture, etc [10]. Different empirical practices have also been done in the field of weather forecasting to obtain accurate results because their high accuracy and reliability in analyzing dataset patterns are exceptional [11], [12].

Machine learning is an artificial intelligence type that can help to make predictions based on new data without needing human help. There are lots of applications which has good models for predicting weather using machine learning(ML) because ML models are capable of finding complex patterns [13], [14]such as classification, regression, and time series analysis. Artificial Intelligence (AI) has largely supplanted the traditional Numerical Weather Prediction (NWP) forecasting approach, which had been followed by Nepal. Various research has been done for predicting the daily, monthly, and annual rainfall prediction by using data mining techniques [15], [16], [17] machine learning algorithms [18], [19] and so many deep learning algorithms and methods [20], [21], [22], [23] as well as several works have been done for air temperature too. Most of the research has been done on predicting the daily [24], [25], [26], and very little research has been done prediction on hourly temperature using machine learning and deep learning techniques [27], [28], [29]. There were several research [20], [29], [30] that concluded that machine and deep learning to predict air temperature or precipitation by using sequential or time series data, deep learning, particularly Recurrent Neural Network type Long short-term memory (LSTM) gives the more precise and accurate result. On the basis of these research findings, we use LSTM to model the forecasting precipitation and temperature among the different stations. As much as co-variate parameters are available, the result is significantly improved [31]. Weather forecasting maintains the quality of life by mitigating the economic crisis and promoting better public health. The safety and well-being of humanity are highly impactable by weather changes [32]. RNNs are explored for meteorological time series [33] and use feedback connections that enable them to retain data that is previously fetched into their architecture. The architecture of RNN has a limitation of its inability to learn and make long-term forecasts [34]. LSTM is a type of ANN with memory cells that control the flow of information into and out of its cells, which have been created to overcome the limitations of RNN [35]. Paper [36] suggests that LSTM is superior to other neural networks for multi-step ahead predictions.

The short term weather forecasting, often called nowcasting, which is very crucial for decision-making in the field of weather. Nowcasting informs operations has a wide horizon of applications, including emergency response, energy management, flood-warning systems, air traffic control, and so many others [37]. Ensemble numerical weather prediction (NWP) systems, which are very computationally extensive, simulate the physical equations of the atmospheric parameters to generate the forecastings, but when the time is shorter for nowcasting, the accuracy of that result tends to be poor [38], [39]. As a result, alternative methods that have good command in making predictions are needed ans, especially precipitation forecasting is based on radar data with high spatial and temporal resolution. But, in those places where radar data is not available then, relying on gauge station data is the only option. Weather forecasting has now entered the era of big data to make the system more advanced, and the volume of ground data makes more robust predictions by adopting deep-learning-based techniques rather than traditional computational intelligence [40]. The main objective of this research is to fit the parameters into the LSTM model and, with the help of this model, forecast the precipitation and air temperature of the stations using time series forecasting of the deep learning approach.

2. Study Area

Kaski is located at latitude 28◦18′19″ N and 84◦4′37″ E, with an altitude that varies from the lowest land range of 450 meters to the highest Himalayan range of 8091 meters [41]. Pokhara is the administrative headquarters of the Kaski district, which covers an area of 2,017 square km. In general, a lot of rain falls from May to September, among which the wettest month is July and the driest month is November, with 402 mm (15.8 inches) and 9 mm (0.4 inches) of precipitation, respectively, whereas the annual average precipitation of Kaski is 1620 mm (63.8 inches). Similarly, the average annual Maximum and minimum precipitation ranges between 20° Celsius and 7° Celsius, with June being the warmest month, with 25° Celsius on average, and January being the coolest month, with 12° Celsius on average [42]. The study stations are visualized in Figure.

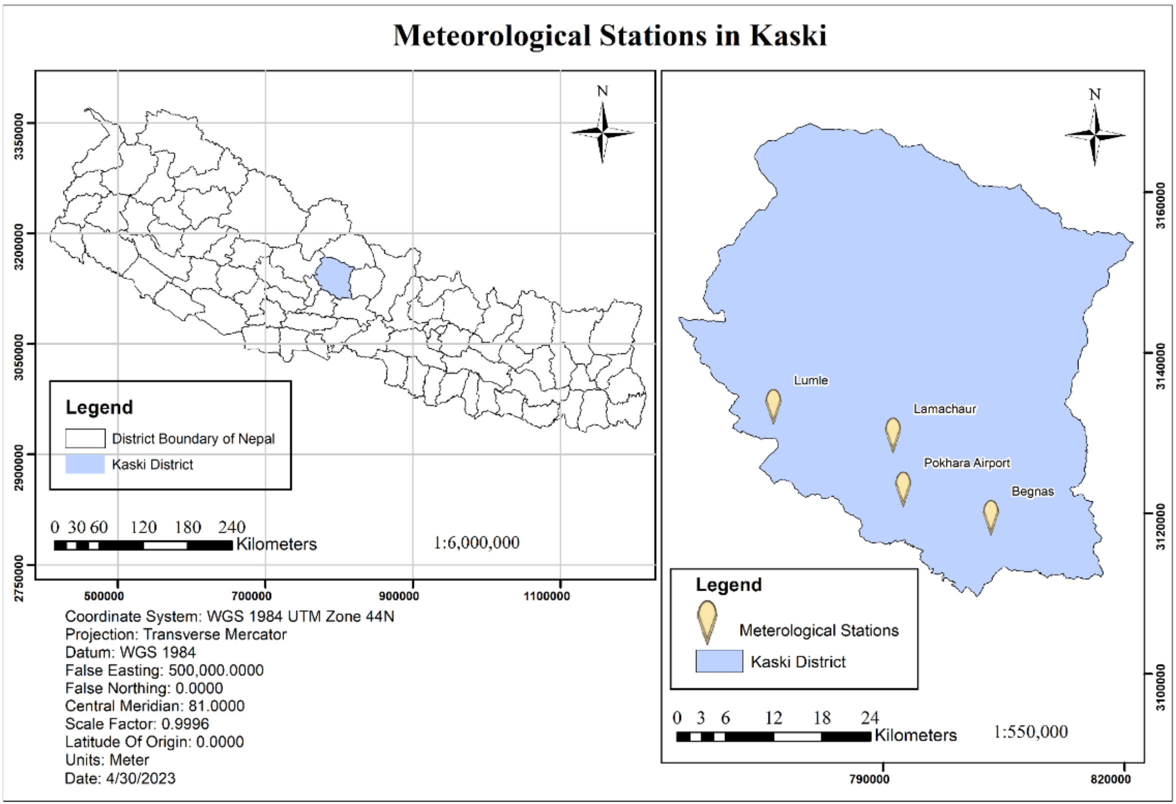

Figure 1: Study Area Map of Meteorological Stations in Kaski District.

Geographic coordinates of the meteorological stations of the study area are shown in Table 1

Table 1: Stations Geographic Details.

SN	Stations Name	District	Latitude	Longitude	Elevation(m)
1	Pokhara Domestic Airport	Kaski	28.20	83.97	827
2	Lumle Station	Kaski	28.29	83.81	1738
3	Begnas Station	Kaski	28.16	84.08	682
4	Lamachaur Station	Kaski	28.26	83.96	991

3. Methodology

The basic workflow for this project is collection of dataset, preprocess this dataset to make applicable to feed into the model and fit this dataset into the model with different layers and hyperparameters. The Figure 2 gives the figurative insights of this project workflow.

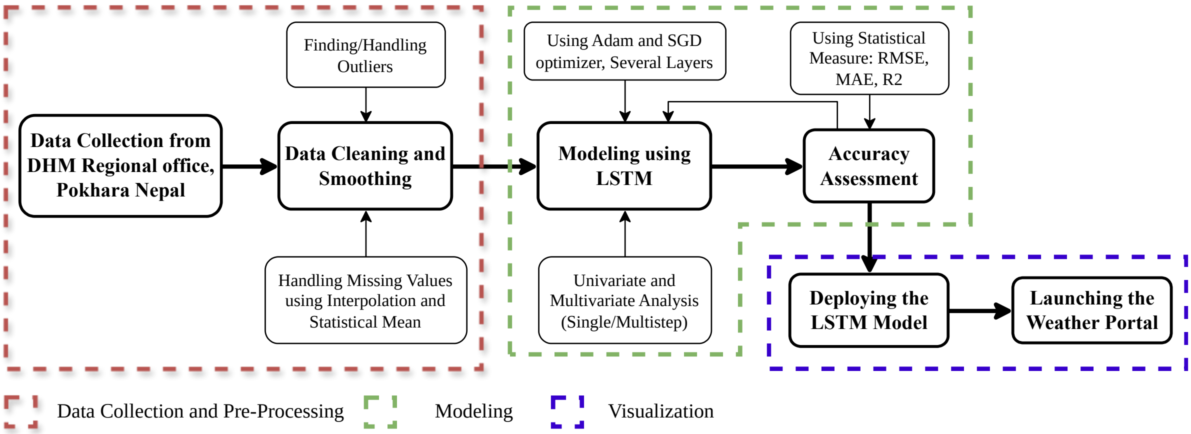

Figure 2: The methodology followed in this research.

3.1 Data Collection

The past weather dataset of four stations was collected from the meteorological regional office in Pokhara, Kaski. Pokhara Domestic Airport only has a dataset of hourly temporal resolution, and the rest of the stations were limited to the daily dataset. The entire dataset used in this research is mentioned in Table 2. Here, the station types, the period that we take for the modeling, and information on the parameters of the respective station were clearly mentioned.

Table 2: Descriptions of Dataset of Different Meteorological Stations used in this study.

SN	Stations Name	Station Type	Frequency	Parameter	Period	No. of Dataset
1	Pokhara Airport	AeroSynoptic	Hourly	Precipitation(mm), Air Temperature (d. C), R.H(%), Wind Speed(m/s)	From 2019-11-11 To 2023-04-16	30051
2	Lumle Station	Agro Meteoro-logical	Daily	Precipitation(mm), Max. and Min. Temperature (d. C), R.H(%), Wind Speed (Knot)	From 2010-01-01 To 2023-04-16	4852
3	Begnas Station	Climatological	Daily	Precipitation(mm), Max and Min Temperature (d. C), R.H(%)	From 2010-01-01 To 2022-12-31	4749
4	Lamachaur Station	Precipitation	Daily	Precipitation(mm)	From 2010-01-01 To 2023-04-16	4852

The sample dataset which had used for the modeling to forecast was mentioned in Table 3 and

Table 4.

Table 3: Sample Hourly data of Pokhara Domestic Airport Station.

Time Stamp	Precipitation (mm)	Air Temperature (°C)	Relative Humidity (%)	Wind speed (m/s)
11/11/2019 6:00	0	23.4	69.399	2.4
11/11/2019 7:00	0	24.3	64.7	2.5

Table 4: Sample daily dataset of Lumle, Begnas and Lamachaur Meteorological Stations.

Time Stamp	Sample Data of Lumle and Begnas Station				Lamachaur
	Precipitation (mm)	Max Temperature (°C)	Min Temperature (°C)	Relative Humidity (%)	Precipitation (mm)
01/01/2015 3:00	0	15	8	32.9	0.0
02/01/2015 3:00	29	12.5	7	95.9	19.6

All the timestamps mentioned in the dataset are in UTC (GMT+5:45) format.

3.2 Data Preprocessing

The original dataset contained 4.85% of missing data (precipitation 2.289290%, air temperature 1.087413%, relative humidity 1.083597%, wind speed 1.09122%) in the hourly data set. Similarly, in Begnas station 4.28% of data was found missing, and in Lumle and Lamachaur station contain few number of missing data. The outliers were detected using the boxplot, and replace these outliers by using pandas with the threshold value, which was assumed by analyzing the boxplot. The threshold is set in this preprocessing to remove the peak values of the parameters because one extremely high or low value has a significant influence on the model. Outliers can sometimes occur while reporting gauge readings manually, and even sometimes, there might be extreme weather events, sensor malfunction, siting issues, low-cost sensors, etc, which significantly impact modeling [43]. Linear interpolation is applied for missing value treatment whereas in case of missing precipitation, fill with zero. The comparison between before and after removing noise data using a box plot of Pokhara airport data as a sample is shown in Figure 3. In Figure 3 (a), the dataset contains noises, and this affects the outcome, so we need to omit this dataset. So, as a result, 3(b), which was created by applying a threshold, contains no noise data.

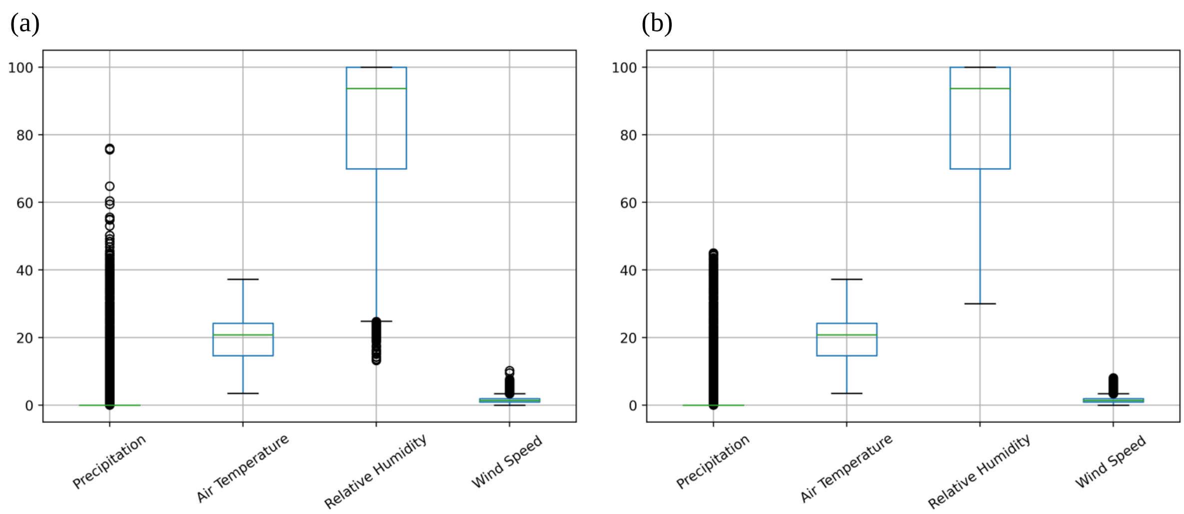

Figure 3: (a) contains the noise and (b) contains the without noise data of Pokhara Domestic Airport Hourly Data frame.

The Pearson correlation matrix was analyzed among the different parameters so that this will provide more in-depth insights into the positive and negative correlation between them, which is very crucial for forecasting.

Figure 4(a) shows the impact between wind speed and relative humidity and air temperature with relative humidity are negatively correlated, whereas for daily station parameters, i.e.,

Figure 4 (b) and

Figure 4 (c), minimum temperature and relative humidity influences have negative relation and rest of the parameters have positive correlation. Paper [44] also concludes that windspeed and minimum temperature have a significant impact on forecasting rainfall.

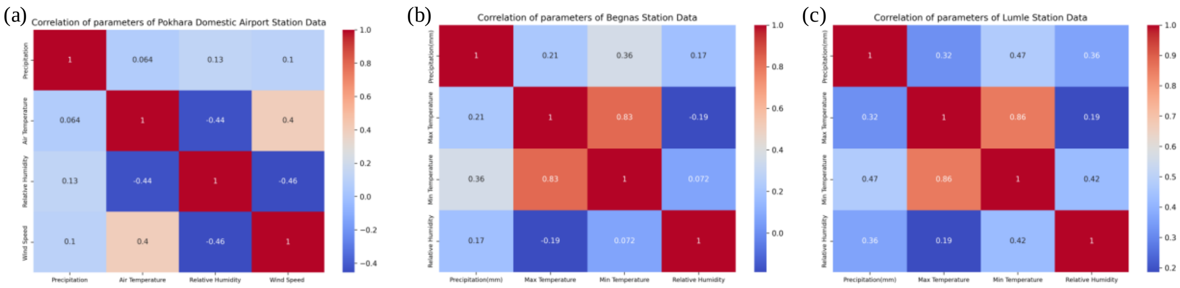

Figure 4: Correlations between the parameters in respective meteorological stations containing several parameters.

3.3 Modeling using Long Short-Term Memory (LSTM)

3.3.1 Architecture of Model

Long Short-Term Memory (LSTM) is an artificial neural network used in deep learning. Artificial Neural Network (ANN) is used for forecasting because of its versatility and capabilities based on past knowledge [45]. LSTM includes the layer of gates (the cell state eq.4 is managed by the input gate eq.1 and forget gate eq.2, which is long term memory and the output gate eq.3 produces the output vector eq.5, which is the memory system that enables to remember a long time) that allows the passing of data through a multi-step process to enable the recognition of patterns [46] which can be seen clearly in the basic architecture of the model in figure 5.

(1) $$i_t = \sigma(W_{ix}x_t + U_{ih}h_{t-1} + b_i)$$

(2) $$f_t = \sigma\left(W_{fx}x_t + U_{fh}h_{t-1} + b_f\right)$$

(3) $$c_t = f_t \cdot c_{t-1} + i_t \cdot \tanh(W_{cx}x_t + U_{ch}h_{t-1} + b_c)$$

(4) $$o_t = \sigma(W_{ox}x_t + U_{oh}h_{t-1} + b_o)$$

(5) $$h_t = o_t \cdot \tanh(c_t)$$

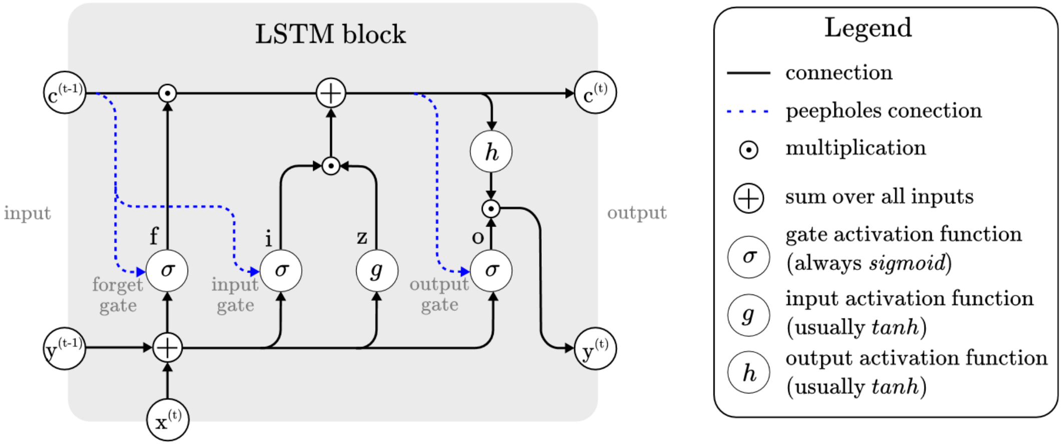

Figure 5: The architecture of the normal LSTM model [47]

The hourly precipitation and average air temperature for Pokhara Airport station were modeled using multivariate multistep LSTM, while daily precipitation, minimum and maximum air temperature for Lumle and Begnas station were modeled using multivariate LSTM and the daily precipitation of Lamachaur station was modeled using univariate LSTM. All the datasets were normalized using the respective scaler as shown in Table, and the training testing dataset was split in the ratio of 80:20 percentage, which is 24021 and 6005 training and testing datasets, respectively, for Pokhara airport hourly data whereas a 70:30 percentage ratio was taken for rest of the daily data stations, i.e., 3396 and 1456 training and testing datasets respectively

Table 5: Final modeling details of all station weather models.

Models	Trainable Parameters	Normalization Scaler	Layers	Optimizer
Pokhara Airport Hourly Temperature	332,034	StandardScaler	Bidirectio-nal LSTM	Adam (learning rate=.001)
Pokhara Airport Hourly Precipitation	1,208,641	MinMaxScaler	LSTM	SGD (momentum=0.95)
Lumle Minimum Air Temperature	1,130,657	MinMaxScaler	LSTM	SGD (momentum=0.9)
Lumle Maximum Air Temperature	1,130,657	MinMaxScaler	LSTM	SGD (momentum=0.9)
Lumle Precipitation	89,249	MinMaxScaler	LSTM	SGD (momentum=0.9)
Begnas Minimum Air Temperature	89,249	MinMaxScaler	LSTM	SGD (momentum=0.95)
Begnas Maximum Air Temperature	89,249	MinMaxScaler	LSTM	SGD (momentum=0.95)
Begnas Daily Precipitation	30,881	MinMaxScaler	LSTM	SGD (momentum=0.85)

Lamachaur Precipitation	88,225	MinMaxScaler	LSTM	SGD (momentum=0.8)

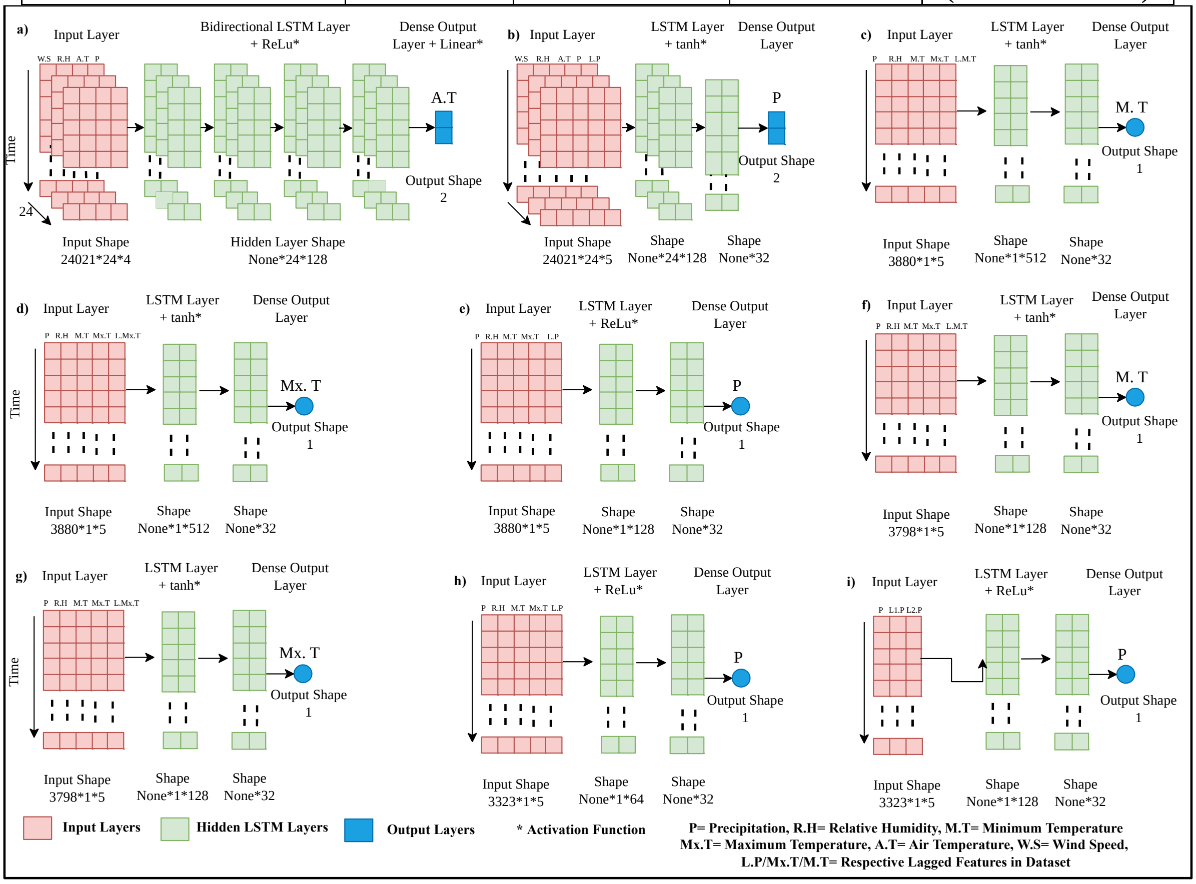

Figure 6: LSTM architecture of different nine models a) Pokhara airport hourly temperature model, b) Pokhara airport hourly temperature model, c) Lumle minimum temperature model, d) Lumle maximum temperature model, e) Lumle Precipitation model, f) Begnas minimum temperature model, g) Begnas maximum temperature model, h) Begnas precipitation model and i) Lamachaur precipitation model

In the Pokhara airport station, hourly modeling sliding window techniques were followed as it contains a large number of the dataset and passed 24 sets of data at once, which contains all the parameters in the normalized form to forecast the precipitation and air temperature for the next 2 hours whereas in another daily forecast, the lag feature is introduced to the data frame as a new column which shifts one-day target data to the future and trains the model with single day dataset to forecast the next day value. Lag features are very inappropriate for processing temporal information like time series forecasting [48], and it is the values of the previous time steps that will be valuable because it is based on the fact that what happened in the past might impact or be inherent the information to the future. Bidirectional LSTM was used for Pokhara airport hourly temperature modeling with Adam optimizer, and the rest of the uses stacked LSTM with Stochastic Gradient Decent (SGD) optimizer. Table 5 contains all the details of the prepared nine models of precipitation and air temperature at four stations.

The complete model architecture of all four stations with all layer properties such as the shape of input layers for LSTM model, numbers and shape of hidden layers with a number of neuron details, activation function used in the model along with the output layer is shown in figure 6.

Figure 7 shows the performance of hourly air temperature and precipitation models on training and testing data. Similarly figure 8, figure 9 and figure 10 indicates the model performance on Lamachaur precipitation model, Lumle and Begnas maximum, minimum air temperature and precipitation model respectively. Overall the prediction value are pretty close to the original one so the performance of the model is good.

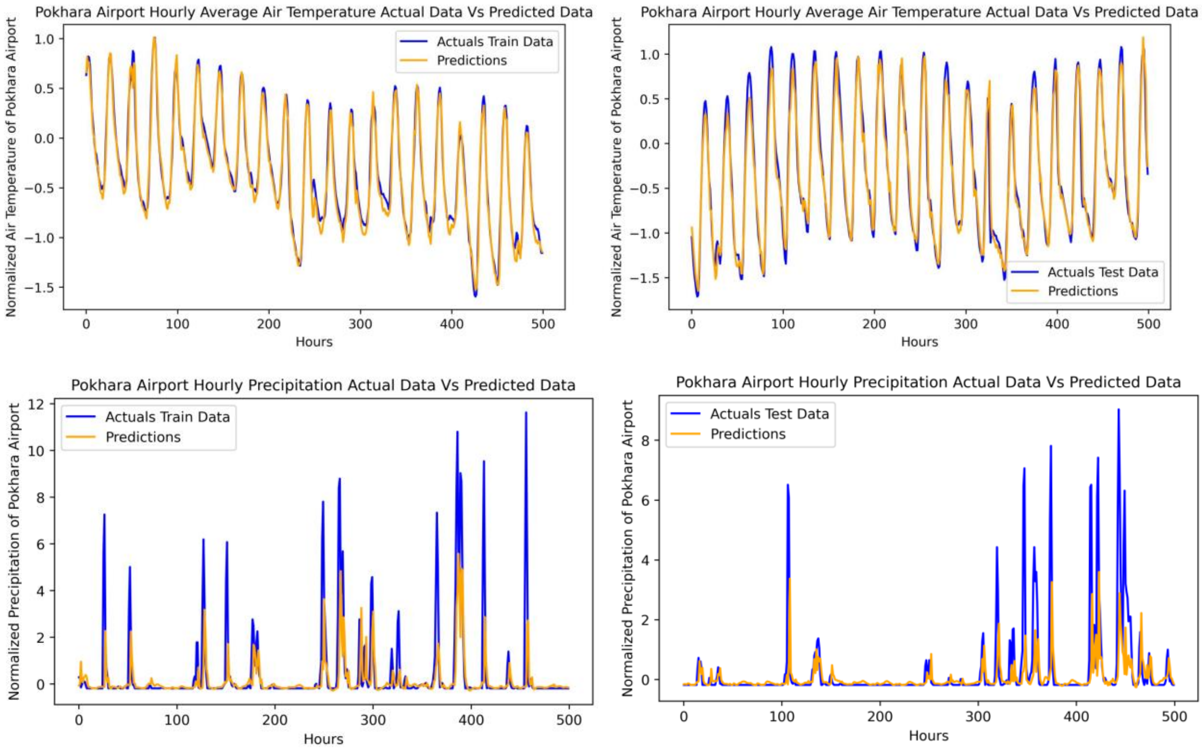

Figure 7: Air temperature and precipitation model performance of Pokhara domestic airport on training and testing data.

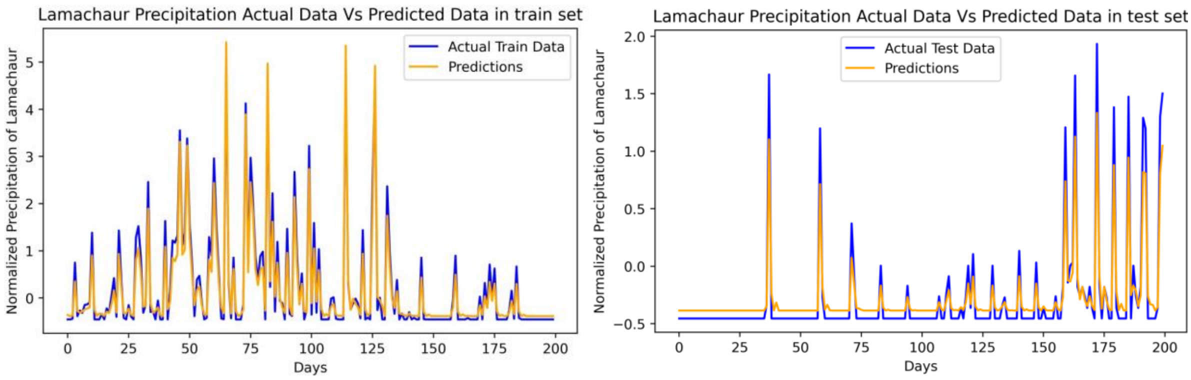

Figure 8: Precipitation model performance of Lamachaur on training and testing data.

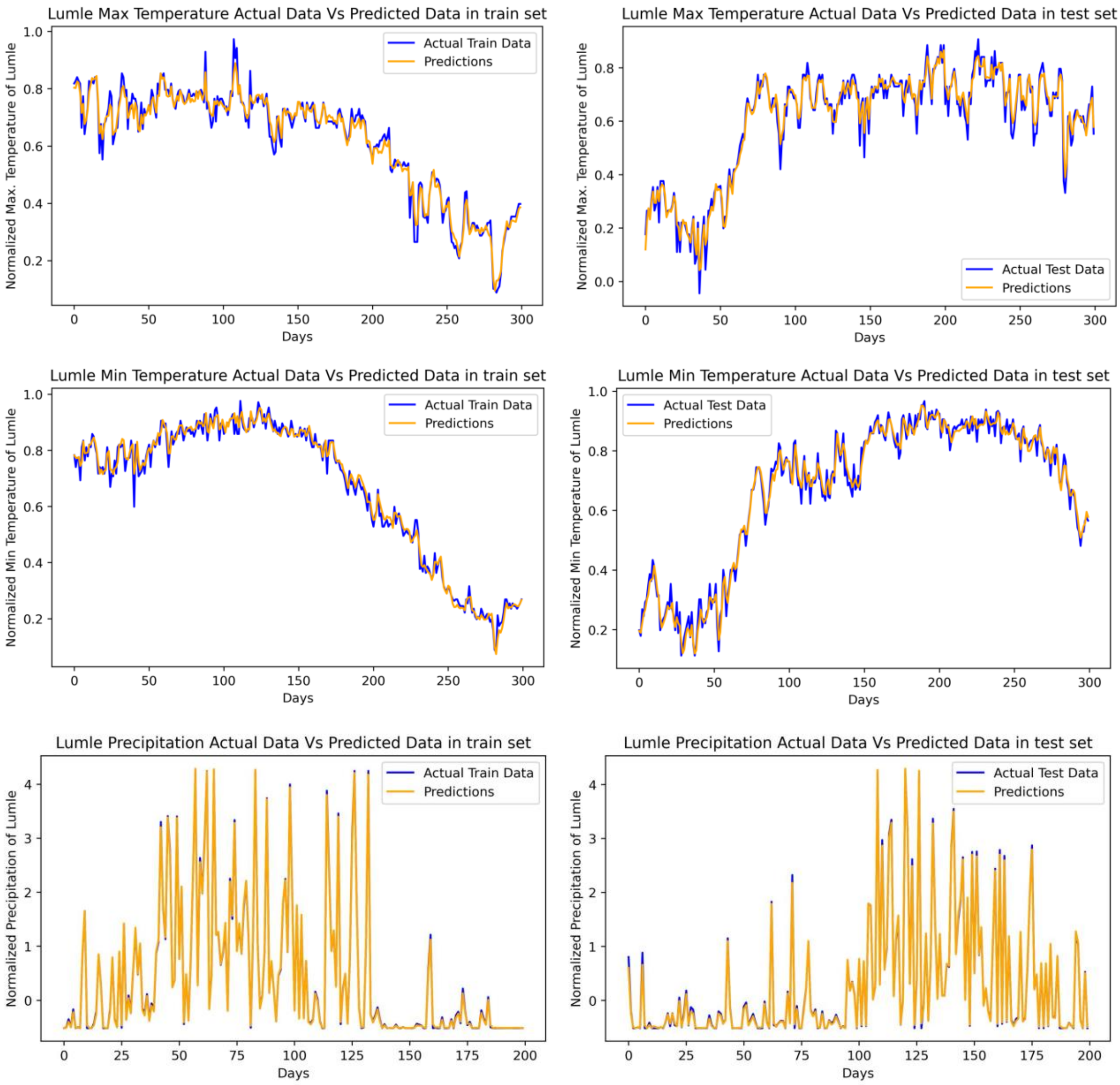

Figure 9: Maximum and minimum air temperature and precipitation model performance of Lumle station on training and testing data.

3.3.1 Evaluation of Model Performance

The Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the R-squared metrics were used to evaluate the performance of the model according to the predicted and measured values from the LSTM model. The square root of average squared differences between actual and predicted observation is RMSE. MAE means average absolute errors between actual and predicted values, whereas R squared measures the extent of variance that how the independent variable of the model is able to relate to the dependent one.

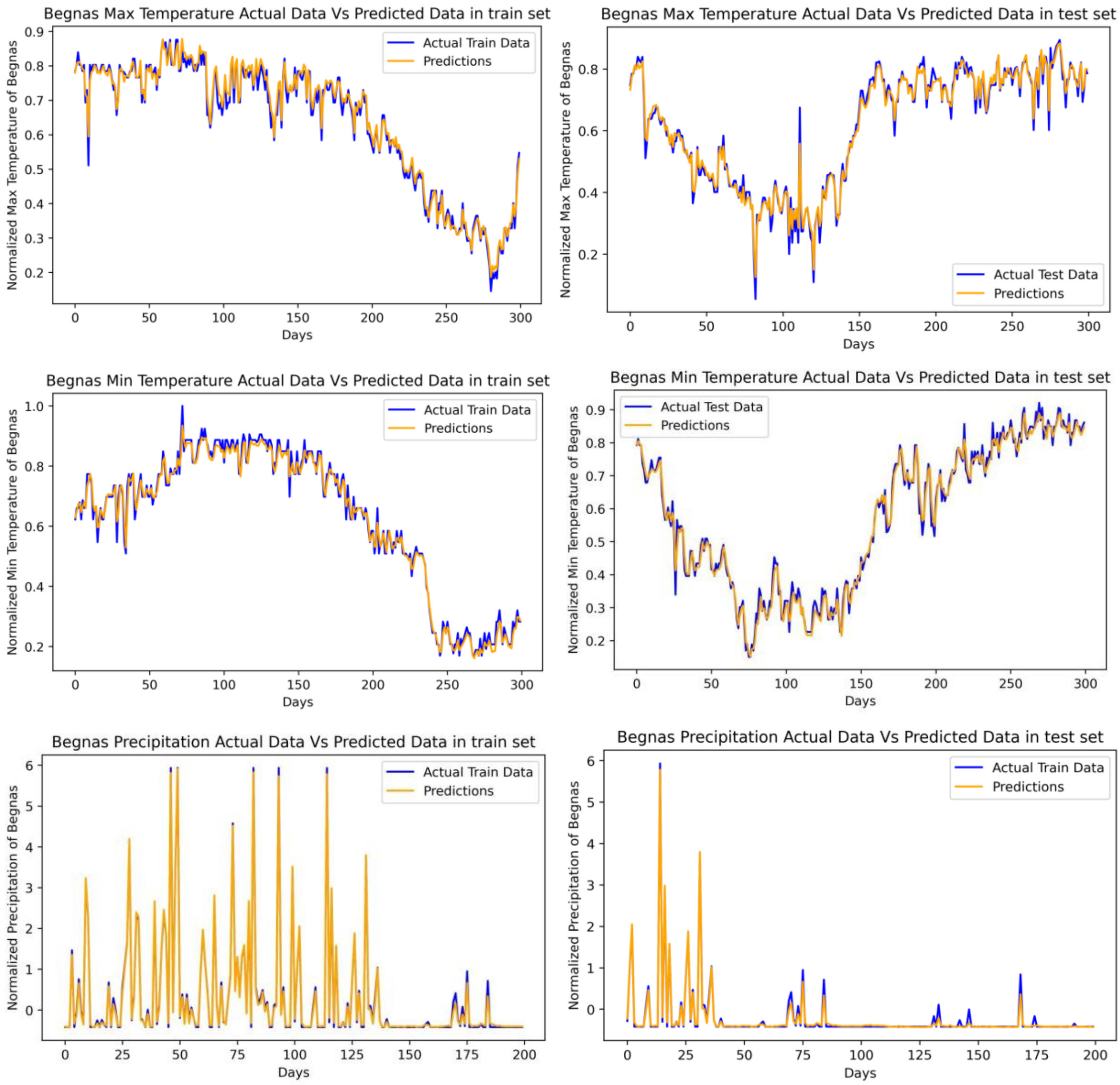

Figure 10: Maximum and minimum air temperature and precipitation model performance of Begnas station on training and testing data.

Adam optimizer was used for the Pokhara airport hourly temperature model and for rest of the model Stochastic Gradient Decent (SGD) was used with different momentum value which is mentioned on table 5.

(6) $$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(a_j - b_j\right)^2}$$

(7) $$\text{MAE} = \frac{1}{n}\sum_{j=1}^{n}\left|a_j - b_j\right|$$

(8) $$R^2 = 1 - \frac{\text{Sum of Squares of Residuals}}{\text{Total Sum of Squares}}$$

3.4 Deploying the Model

A combination of programs and frameworks, including Flask for the backend and React JS for the frontend, have been used to develop a weather portal. Users can access the results by using the web application. The final model is downloaded in hierarchical data format (h5) and they were then loaded into the Flask server along with each station's observational data in CSV file format. The necessary variables and lags were taken from the observational data and stored in a CSV file to prepare the data for prediction. The prediction data was standardized using the respective scaler mentioned in Table. After normalizing the input data, it was fed to the model to provide normalized output, which was then inversely converted to produce denormalized findings. Users were then able to simply receive information about the projected temperature and precipitation for their area of interest. The web application is static since the database consisting weather parameters used to predict temperature and precipitation is limited for a given timestamp. However, it can be made dynamic by feeding the newly observed data from the meteorological stations to the database, which either can be done by manually editing to the CSV file or by pegging the database with the official DHM's data, the latter being more systematic.

4. Findings and Discussion

The hourly precipitation and air temperature for Pokhara Airport station were modeled, while the daily precipitation, minimum and maximum air temperature for Lumle, Begnas station, and daily precipitation of Lamachaur station were modeled using LSTM. The prediction performance evaluation metrics for the modeled LSTM algorithm, RMSE, and MAE were defined, and the R squared value was defined to evaluate the overall fit of the data into the model. Table 6 represents the predictive power of different models in terms of RMSE and MAE for both train and test data, along with the R-squared value for each model's train and test data.

Among the models, precipitation of Lamachaur station had a high value of Root Mean Square Error (RMSE) of 4.08 and Mean Absolute Error (MAE) of 2.87 on test data because it was modeled with only one parameter, i.e. precipitation, followed by Lumle and Begnas precipitation model due to high variation of precipitation patterns. The complex nature of precipitation and its dependencies on a variety of factors plays a significant role in weak predictive power, because of which high magnitude of differences between actual and predicted values were observed in those models when compared to others. In contrast, the hourly precipitation of Pokhara Domestic Airport had a least RMSE and MAE value of 0.61 and 0.21, respectively, on test data, which indicates the difference in magnitude of the actual and predicted value of that station. The superior model performance was achieved in this case because the number of datasets that were used for Pokhara Domestic Airport was larger numbers than at other stations, due to which the model learned the underlying precipitation patterns of this station more significantly.

Similarly, in terms of predicting temperature, the hourly temperature of Pokhara Domestic Airport outperformed other models with RMSE and MAE values of 0.16 and 0.12. This is due to the fact that temperature, more or less, follows seasonal patterns, making it easier to understand the flow of trends for an algorithm along with the provision of a larger number of datasets to detect seasonal change. The minimum and maximum surface air temperature of Lumle and Begnas stations were predicted with the RMSE scores of 0.58 and 0.82 (Lumle) and 0.58 and 0.66 (Begnas). The MAE scores were 0.46 and 0.64 for Lumle and 0.36 and 0.47 for Begnas station.

The measure of R squared value describes fit rather than forecast accuracy; all the models fit to the model very accurately with the actual data. The closer its value to 1 means it's performing better. The values range from 0.89 to 0.99, indicating that the relationships between input and target variables were captured accurately, maintaining a good fit into the model.

Table 6: Train and Test RMSE and MAE of all models.

Models	RMSE		MAE		R Squared	
	Train	Test	Train	Test	Train	Test
Airport Hourly Temperature	0.12	0.16	0.077	0.12	0.98	0.96
Airport Hourly Precipitation	0.78	0.61	0.27	0.21	0.92	0.89
Lumle Minimum Temperature	0.62	0.58	0.47	0.46	0.98	0.99
Lumle Maximum Temperature	0.79	0.82	0.6	0.64	0.96	0.96
Lumle Precipitation	2.16	2.34	1.36	1.52	0.99	0.98
Begnas Minimum Temperature	0.57	0.58	0.42	0.36	0.98	0.99
Begnas Maximum Temperature	0.69	0.66	0.51	0.47	0.98	0.98
Begnas Daily Precipitation	1.55	1.81	0.99	1.15	0.99	0.99
Lamachaur Precipitation	3.78	4.08	2.73	2.87	0.97	0.97

Analyzing the results from Table 6, it is seen that errors in predictions of temperature are relatively lower than in predictions of precipitation of the same meteorological station. This is due to the fact that temperature follows seasonal patterns that can be easily understood by deep learning algorithms, while precipitation comprises a complex nature with more interdependencies parameters such as wind speed, wind direction, atmospheric pressure, etc. This makes the accurate prediction of precipitation more challenging. However, the accuracy can be increased provided that most of the influencing factors for precipitation are taken into account during the data collection process. Unfortunately, the variables recorded in the stations of interest by the DHM did not include a variety of factors responsible for rainfall, which eventually became the shortcoming of the project.

Table 7: Actual VS Predicted values of Pokhara Domestic Airport Station

Model	Actual	Predicted	Timestamp
Pokhara Airport Hourly Precipitation	0	0.031	2023-04-16 7:00
Pokhara Airport Hourly Precipitation	0	0.01	2023-04-16 8:00
Pokhara Airport Hourly Temperature	30	29.5	2023-04-16 7:00
Pokhara Airport Hourly Temperature	30.2	29.8	2023-04-16 8:00

Table 8: Actual VS Predicted Values of Daily Station Models

Model	Actual	Predicted	Date
Begnas Daily Precipitation	0	0.078	2022-12-31
Begnas Daily Maximum Temperature	19	18.65	2022-12-31
Begnas Daily Minimum Temperature	8.5	7.9	2022-12-31
Lumle Daily precipitation	0	0.043	2023-04-14
Lumle Daily Maximum Temperature	26	25.4	2023-04-14
Lumle Daily Minimum Temperature	14.2	13.4	2023-04-14
Lamachaur Daily Precipitation	0	1.46	2023-04-16

We have predicted the air temperature and precipitation of the stations using the final model and compared them with the actual observed data of that predicted date as shown in Table 7 (multi-step forecasting of the next 2 hours) and Table 8 (single-step forecasting). It provides specific knowledge about the accuracy of the proposed models as the numeric values of predicted output can be compared with the actual observed value at the station. As our objective is to predict the next 2-hour forecast of air temperature and precipitation for Pokhara domestic airport meteorological station and the next one-day forecast for Lumle, Begnas, and Lamachaur meteorological stations, it was thus achieved an acceptable result.

5. Conclusion

The precipitation and temperature of all four stations have been modeled using LSTM with different numbers of hidden layers, neurons, and optimizers, as well as the best-suited activation function. The results of the project indicate 10 that the accuracy of the machine learning models can vary significantly depending on the quality and quantity of the datasets and the parameters or variables used in the model. Among the stations, Lamachaur station only contains precipitation parameters; on the basis of this single parameter, the model of predicting next-day precipitation data using univariate LSTM has more error, followed by Lumle, Begnas, and Pokhara Domestic Airport. Although all models fit well for training and testing data based on R2 value, Pokhara Airport has a short temporal resolution of the hourly dataset and a high number of datasets, so compared to other station models, particularly Pokhara Airport's precipitation and air temperature model performs very well in terms of error analysis and all the outputs are deployed through the weather portal. Rather than using complex and tedious Numerical Weather Prediction (NWP), the Machine Learning approach will be the best alternative for the short computational time with efficient results. Based on the findings and complete deployment of the project following are the recommendations for increasing the accuracy of the overall project.

1. Integration of additional weather parameters like due points, cloud state, wind direction, atmospheric pressure, and so on makes the prediction more precise.
2. The use of more dataset will capture the long-term dependencies of weather patterns which helps to give better results.
3. Before using all the historical dataset, calibration of raw data will be highly recommended to know about the biases and errors on the original dataset itself after which it will perform well on the model.
4. Incorporating ensemble forecasting will give more precise results.
5. Further research and studies can explore other machine learning algorithms to improve the model accuracy.

6. Author's Contribution

Supath Dhital has done the conceptualization of the project, modeling, discussions, and conclusions, writing the manuscripts, and handling the visualization of graphs. Kapil Lamsal has helped with the web deployment of the weather portal, researching the background introduction, discussions, writing, and screening the paper. Sulav Shrestha has helped by doing the preprocessing tasks on data, exploring the background study, and checking the suitability of the model. Umesh Bhurtyal is the entire project supervisor, so he timely discusses, reviews, edits, and guides during the whole project.

7. Data Availiability

Data, Modeling, and Weather Portal codes can be accessed via the GitHub repository: Modeling, Weather Portal

7. Conflict of Interest

There is no conflict of interest for this paper.

8. Acknowledgment

We would like to thank the Department of Hydrology and Meteorology (DHM), Nepal, for providing us with meteorological past data required for the project and would like to thank the Department of Geomatics Engineering at the Institute of Engineering (IOE), Pashchimanchal Campus for giving us the chance to design and develop the project as the fulfillment of the bachelor's degree completion.

References

[1] R.B. Stull, "An introduction to boundary layer meteorology". Vol. 13. *Springer Science & Business Media*, 2012. Accessed: Mar. 28, 2024.

[2] "Weather and atmosphere | National Oceanic and Atmospheric Administration." Accessed: Mar. 28, 2024. [Online]. Available: https://www.noaa.gov/education/resource-collections/weather-atmosphere

[3] "DHM." Accessed: Mar. 28, 2024. [Online]. Available: https://www.dhm.gov.np/pages/about-us

[4] R. Karki, "Status of automatic weather stations in Nepal and comparison of air temperature and precipitation data between automatic weather station and manual observation," 2010, Accessed: Mar. 28, 2024.

[5] P. A. Kucera *et al.*, "Precipitation from Space: Advancing Earth System Science," *Bulletin of the Meteorological Society*, vol. 94, no. 3, pp. 365–375, Mar. 2013, https://doi.org/10.1175/BAMS-D-11-00171.1

[6] H. Wu *et al.,* "Evaluation of global flood detection using satellite-based rainfall and a hydrologic model," *Journal of Hydrometeorology,* 2012, Accessed: Mar. 28, 2024. https://doi.org/10.1175/JHM-D-11-087.1

[7] B. Branko, "Geopolitics of climate change: a review," *Thermal Science,* vol. 16, no. 3, pp. 629–654, 2012, https://doi.org/10.2298/TSCI120202127B

[8] Z. L. Li *et al.*, "Satellite-derived land surface temperature: Current status and perspectives," *Remote Sensing Environment*, vol. 131, pp. 14–37, Apr. 2013, https://doi.org/10.1016/j.rse.2012.12.008

[9] W. W. Immerzeel, L. P. H. Van Beek, and M. F. P. Bierkens, "Climate change will affect the asian water towers," *Science (1979)*, vol. 328, no. 5984, pp. 1382–1385, Jun. 2010, https://doi.org/10.1126/science.1183188

[10] R. E. Abdel-Aal, "Hourly temperature forecasting using abductive networks," *Engineering Applications of Artificial Intelligence*, vol. 17, no. 5, pp. 543–556, Aug. 2004, https://doi.org/10.1016/j.engappai.2004.04.002

[11] C. Penland, L. Matrosova.-, "Prediction of tropical Atlantic Sea surface temperatures using linear inverse modeling," *Journal of Climate, 1998*, https://doi.org/10.1175/1520-0442(1998)011%3C0483:POTASS%3E2.0.CO;2

[12] C. Penland, T. Magorian, "Prediction of Niño 3 sea surface temperatures using linear inverse modeling," *Journal of Climate*, 1993, Accessed: Mar. 28, 2024. https://doi.org/10.1175/1520-0442(1993)006%3C1067:PONSST%3E2.0.CO;2

[13] K. P. Moustris, K. P. Moustris, I. K. Larissi, P. T. Nastos, and A. G. Paliatsos, "Precipitation forecast using artificial neural networks in specific regions of Greece," *Springer, Water Resources Management,* vol. 25, no. 8, pp. 1979–1993, Jun. 2011, https://doi.org/10.1007/s11269-011-9790-5

[14] B. T. Pham *et al.*, "Development of advanced artificial intelligence models for daily rainfall prediction," *Atmospheric Research*, vol. 237, Jun. 2020, https://doi.org/10.1016/j.atmosres.2020.104845

[15] A. Kusiak, X. Wei, A. P. Verma, and E. Roz, "Modeling and prediction of rainfall using radar reflectivity data: A data-mining approach," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 4, pp. 2337–2342, 2013, https://doi.org/10.1109/TGRS.2012.2210429

[16] K. K. Chowdari, R. Girisha, and K. C. Gouda, "A study of rainfall over India using data mining," *2015 International Conference on Emerging Research in Electronics, Computer Science and Technology, ICERECT 2015*, pp. 44–47, Jun. 2016, https://doi.org/10.1109/ERECT.2015.7498985

[17] V. P. Tharun, R. Prakash, and S. R. Devi, "Prediction of Rainfall Using Data Mining Techniques," *Proceedings of the International Conference on Inventive Communication and Computational Technologies, ICICCT 2018*, pp. 1507–1512, Sep. 2018, https://doi.org/10.1109/ICICCT.2018.8473177

[18] C. Z. Basha, N. Bhavana, P. Bhavya, and V. Sowmya, "Rainfall Prediction using Machine Learning Deep Learning Techniques," *Proceedings of the International Conference on Electronics and Sustainable Communication Systems, ICESC 2020*, pp. 92–97, Jul. 2020, https://doi.org/10.1109/ICESC48915.2020.9155896

[19] C. M. Liyew and H. A. Melese, "Machine learning techniques to predict daily rainfall amount," *Journal of Big Data*, vol. 8, no. 1, pp. 1–11, Dec. 2021, https://doi.org/10.1186/s40537-021-00545-4

[20] P. Hewage, A. Behera, M. Trovati, and E. Pereira, "Long-Short Term Memory for an Effective Short-Term Weather Forecasting Model Using Surface Weather Data," *IFIP International Conference on Artificial Intelligence Applications and Innovations*, vol. 559, pp. 382–390, 2019, https://doi.org/10.1007/978-3-030-19823-7_32

[21] J. Kang *et al.*, "Prediction of Precipitation Based on Recurrent Neural Networks in Jingdezhen, Jiangxi Province, China," *Atmosphere,* vol. 11, no. 3, p. 246, Feb. 2020, https://doi.org/10.3390/atmos11030246

[22] C. J. Gamboa-Villafruela, J. C. Fernández-Alvarez, M. Márquez-Mijares, A. Pérez-Alarcón, and A. J. Batista-Leyva, "Convolutional LSTM Architecture for Precipitation Nowcasting Using Satellite Data," *Environmental Sciences Proceedings 2021, Vol. 8, Page 33*, vol. 8, no. 1, p. 33, Jun. 2021, https://doi.org/10.3390/atmos11030246

[23] A. N. Caseri, L. B. Lima Santos, and S. Stephany, "A convolutional recurrent neural network for strong convective rainfall nowcasting using weather radar data in Southeastern Brazil," *Artificial Intelligence in Geosciences*, vol. 3, pp. 8–13, Dec. 2022, https://doi.org/10.1016/j.aiig.2022.06.001

[24] B. Ustaoglu, H. K. Cigizoglu, and M. Karaca, "Forecast of daily mean, maximum and minimum temperature time series by three artificial neural network methods," *Meteorological Applications*, vol. 15, no. 4, pp. 431–445, Dec. 2008, https://doi.org/10.1002/met.83

[25] M. Murat, I. Malinowska, M. Gos, and J. Krzyszczak, "Forecasting daily meteorological time series using ARIMA and regression models," *International Agrophysics*, vol. 32, pp. 253–264, 2018, http://dx.doi.org/10.1515/intag-2017-0007

[26] M. L. Lin, C. W. Tsai, and C. K. Chen, "Daily maximum temperature forecasting in changing climate using a hybrid of Multi-dimensional Complementary Ensemble Empirical Mode Decomposition and Radial Basis Function Neural Network," *Journal of Hydrology: Regional Studies*, vol. 38, p. 100923, Dec. 2021, https://doi.org/10.1016/j.ejrh.2021.100923

[27] D. Shah, "Short term temperature forecasting using LSTMS, and CNN," 2021, Accessed: Mar. 28, 2024. https://digitalcommons.njit.edu/theses/1840

[28] D. Carrión *et al.*, "A 1-km hourly air-temperature model for 13 northeastern U.S. states using remotely sensed and ground-based measurements," *Environmental Research*, vol. 200, p. 111477, Sep. 2021, https://doi.org/10.1016/j.envres.2021.111477

[29] J. Hou, Y. Wang, J. Zhou, and Q. Tian, “Prediction of hourly air temperature based on CNN–LSTM,” *Geomatics, Natural Hazards and Risk*, vol. 13, no. 1, pp. 1962–1986, Dec. 2022, https://doi.org/10.1080/19475705.2022.2102942

[30] K. Huang, W. Liu, Y. Li, B. Vucetic, “To retransmit or not: Real-time remote estimation in wireless networked control,” *IEEE International Conference on Communications (ICC), 2019*, https://doi.org/10.1109/ICC.2019.8761710

[31] D. Ji *et al.*, “Assessing Parameter Importance of the Weather Research and Forecasting Model Based on Global Sensitivity Analysis Methods,” *Journal of Geophysical Research: Atmospheres*, vol. 123, no. 9, pp. 4443–4460, May 2018, https://doi.org/10.1002/2017JD027348

[32] J. K. Lazo, H. R. Hosterman, J. M. Sprague-Hilderbrand, and J. E. Adkins, “Impact-Based Decision Support Services and the Socioeconomic Impacts of Winter Storms,” *Bulletin of American Meteorology Society*, vol. 101, no. 5, pp. E626–E639, May 2020, https://doi.org/10.1175/BAMS-D-18-0153.1.

[33] M. M. P. Ramos, C. L. Del Alamo, and R. A. Zapana, “Forecasting of Meteorological Weather Time Series Through a Feature Vector Based on Correlation,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11678 LNCS, pp. 542–553, 2019, https://doi.org/10.1007/978-3-030-29888-3_44

[34] L. Ni *et al.*, “Streamflow and rainfall forecasting by two long short-term memory-based models,” *Journal of Hydrology*, vol. 583, p. 124296, Apr. 2020, https://doi.org/10.1016/j.jhydrol.2019.124296

[35] K. Greff, R. K. Srivastava, J. Koutnik, B. R. Steunebrink, and J. Schmidhuber, “LSTM: A Search Space Odyssey,” *IEEE Trans Neural Network and Learning System*, vol. 28, no. 10, pp. 2222–2232, Oct. 2017, https://doi.org/10.1109/TNNLS.2016.2582924

[36] Y. Liu, D. Hou, J. Bao, and Y. Qi, “Multi-step ahead time series forecasting for different data patterns based on LSTM recurrent neural network,” *Proceedings - 2017 14th Web Information Systems and Applications Conference, WISA 2017*, vol. 2018-January, pp. 305–310, Jul. 2017, https://doi.org/10.1109/WISA.2017.25

[37] J. Wilson, Y. Feng, M. Chen, R.D. Roberts, “Nowcasting challenges during the Beijing Olympics: Successes, failures, and implications for future nowcasting systems,” *Weather and Forecasting*, 2010, https://doi.org/10.1175/2010WAF2222417.1

[38] M. Buehner, D. Jacques, “Non-Gaussian deterministic assimilation of radar-derived precipitation accumulations,” *Monthly Weather Review*, 2020, https://doi.org/10.1175/MWR-D-19-0199.1

[39] J. Sun, “Convective-scale assimilation of radar data: progress and challenges,” *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography,* 131, no. 613, pp. 3439–3463, Jan. 2005, https://doi.org/10.1256/qj.05.149

[40] K. U. Jaseena and B. C. Kovoor, “Deterministic weather forecasting models based on intelligent predictors: A survey,” *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 6, pp. 3393–3412, Jun. 2022, https://doi.org/10.1016/j.jksuci.2020.09.009

[41] K. Basnet, A. Shrestha, P. C. Joshi, and N. Pokharel, “Analysis of Climate Change Trend in the Lower Kaski District of Nepal,” *Himalayan Journal of Applied Science and Engineering*, vol. 1, no. 1, pp. 11–22, Dec. 2020, https://doi.org/10.3126/hijase.v1i1.33536

[42] “World weather, climate and accurate forecast information.” Accessed: Mar. 28, 2024. [Online]. Available: https://weather-and-climate.com/

[43] L. Ma, X. Gu, and B. Wang, "Correction of Outliers in Temperature Time Series Based on Sliding Window Prediction in Meteorological Sensor Network," *Information*, vol. 8, no. 2, p. 60, May 2017, https://doi.org/10.3390/info8020060

[44] M. S. Pathan, J. Wu, Y. H. Lee, J. Yan, and S. Dev, "Analyzing the Impact of Meteorological Parameters on Rainfall Prediction," *2021 IEEE USNC-URSI Radio Science Meeting (Joint with AP-S Symposium), USNC-URSI 2021 - Proceedings*, pp. 100–101, 2021, https://doi.org/10.23919/USNC-URSI51813.2021.9703664

[45] S. Siami-Namini, N. Tavakoli, and A. Siami Namin, "A Comparison of ARIMA and LSTM in Forecasting Time Series," *Proceedings - 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018*, pp. 1394–1401, Jul. 2018, https://doi.org/10.1109/ICMLA.2018.00227

[46] H. Hewamalage, C. Bergmeir, and K. Bandara, "Recurrent Neural Networks for Time Series Forecasting: Current status and future directions," *International Journal of Forecasting*, vol. 37, no. 1, pp. 388–427, Jan. 2021, https://doi.org/10.1016/j.ijforecast.2020.06.008

[47] G. Van Houdt, C. Mosquera, and G. Nápoles, "A review on the long short-term memory model," *Artificial Intelligence Review*, vol. 53, no. 8, pp. 5929–5955, Dec. 2020, https://doi.org/10.1007/s10462-020-09838-1

[48] M. Taghi Sattari, K. Yurekli, and M. Pal, "Performance evaluation of artificial neural network approaches in forecasting reservoir inflow," *Applied Mathematical Modelling*, vol. 36, no. 6, pp. 2649–2657, Jun. 2012, https://doi.org/10.1016/j.apm.2011.09.048