DOI: https://doi.org/10.1007/978-981-99-6702-5_14

On the Impact of Temperature for Precipitation Analysis

Madara Premawardhana, Menatallah Abdel Azeem, Sandeep Singh Sengar, Soumyabrata Dev

Abstract Climate is the result of the constant interaction between different weather variables where Temperature and Precipitation are significant factors. Precipitation refers to the condensation of water vapor from clouds as a result of gravitational pull. These variables act as governing factors for determining rainfall, snowfall, and air pressure while determining wide-ranging effects on ecosystems. Different calculation methods could be employed such as Standard Precipitation Index for determining precipitation. Temperature is the measure that is used to identify the heat energy generated by solar radiation and other industrial factors. For understanding the interplay between these two variables, data gathered from several regions of the world including North America, Europe, Australia, and Central Asia were analyzed and the findings are presented in this paper. Prediction methods such as Multiple Linear Regression and LSTM (Long-Short-Term-Memory) have been employed for predicting rainfall from Temperature and Precipitation data. The inter-dependency of other weather parameters are also observed in this paper relating to rainfall prediction. The accuracy of the prediction models using machine learning has also been experimented within the study. The implementation of our work is available here.

1 Introduction

Climatic changes occur due to the interplay between different weather variables. Temperature and precipitation could be identified as two major weather factors which are strongly dependent on each other [1]. While all weather and meteorological variables affect rainfall classification greatly, only a few of them have greater influence such as solar radiation [2], Precipitable Water Vapor (PVW) [3], seasonal

Madara Premawardhana
School of Computing, University of Buckingham, United Kingdom, e-mail: 2205279@buckingham.ac.uk

Menatallah Abdel Azeem
University College Dublin, Ireland, e-mail: menatalla.abdelazim@ucdconnect.ie

Sandeep Singh Sengar
Department of Computer Science, Cardiff Metropolitan University, United Kingdom, e-mail: SSSengar@cardiffmet.ac.uk

Soumyabrata Dev
School of Computer Science, University College Dublin, Ireland e-mail: soumyabrata.dev@ucd.ie

and diurnal features [4]. Earth receives temperature from solar radiance. Human activities such as rapid industrialization contribute to the global heat increase. The precipitation is water that comes down from the sky as rain, snow, hail, or any other form [5]. Some places on Earth which have very high precipitation, experience very wet climates, and some have low precipitation, resulting in very dry climates. The interplay between temperature and precipitation also differs on factors such as Earth's natural orientation. Rainfall prediction and its accuracy is important for long-term decision-making in a variety of areas including agriculture, irrigation, and scientific research [6]. Observing the prediction of rainfall occurrence by analyzing temperature and precipitation, we understand that each factor's independent nature is important. The variation of each variable being influenced by other weather factors also has to be taken into consideration when analyzing the relationship between the selected factors. Furthermore, we have allocated data sets from around the world to cover different climatic regions such as North America, Europe, Central Asia, Canada, and Australia which are further discussed in Section 2. The data sets were collected from National Snow and Ice Data Center (NSDIC) and the National Oceanic and Atmospheric Administration (NOAA) which possess accurate weather data sets in a standard format [7]. These data sets were analyzed for anomalies and non-available values and accordingly were cleaned. Then correlation analysis was done on the available data for identifying the relationship between different weather variables including temperature and precipitation. The selected weather variables were also predicted using the simple machine-learning technique of multiple linear regression by using the available data. The potentiality of predictions of temperature and precipitation is also presented in the paper. Further, employing the weather data including temperature and precipitation, multiple linear regression and Long Short-Term Memory (LSTM) models were developed to predict the rainfall in the Sydney-Australia region.

The main contributions of the paper are as follows:

- Analyzing and identifying the relationship between different environmental factors towards the prediction of proportionately collaborating environmental factors using Simple Linear Regression (SLR) and Long Short-Term Memory (LSTM). Here , the focus is on comparing the results of the two aforementioned methodologies for identifying the most suitable approach.
- Prediction of rainfall depended on using independent data sets which contain temperature and precipitation from different areas of the world and adopting base work for accurate rainfall prediction using Artificial Neural Networks. This procedure could be easily re-modified and re-applied for finding relationships and predicting rainfall using a variety of other variables as well.
- In order to support reproducible and extensible research, the code to reproduce the results of this manuscript is available online[1].

[1] The source code related to this paper is available via `https://github.com/MadaraPremawardhana/On-the-Impact-of-Temperature-for-Precipitation-Analysis`.

2 Data Description

The data sets utilized for the prediction and correlation analysis carried out in the paper, were taken from the National Oceanic and Atmospheric Administration (NOAA), National Snow and Ice Data Center (NSIC) [8]. The data taken from NOAA includes three locations for the duration of 5 years in a temporal resolution of 24 hours. The data collection locations are Alpena-USA, Beaucelville-Canada, and Dublin-Ireland. The data sets include weather data such as Snow Depth, Average Temperature, Precipitation, Snow, the direction of a peak wind gust, peak gust wind speed, Minimum Temperature, Maximum Temperature, Average Wind Speed, the direction of the fastest 2-minute wind, the fastest 2-minute wind speed, the direction of the fastest 5-minute wind, and the fastest 5-second wind speed which were represented by SNWD, TAVG, PRCP, SNOW, WSFG, WSF5, TMIN, TMAX, AWND, WDF2, WSF2, WDF5 and WSF5 respectively [9].

The data taken from NSIC includes data from Central Asia, including data from Almaty, Kazakhstan from 1879-2003. This data has not been employed in carrying out the predictions or analyzing correlations since the data set had a temporal resolution of 1 month, including data on Average Temperature and Average Precipitation only. Hence, this data has been used for identifying precipitation variation and temperature in Section 5 before carrying out correlation analysis for the data sets of Dublin, Alpena, and Beauceville. Furthermore, data sets of Lake Michigan Riverfront, Chicago-USA, and Sydney-Australia have been taken from NOAA and Kaggle respectively. The Sydney, Australia data set contains Minimum and Maximum Temperatures, Rainfall, Wind Gusts, Wind Gust Direction, Humidity, Wind Speed, Pressure, Cloud density, and Rain status of the day and the day after. Chicago, USA data set contains weather variables: Rain Interval, Precipitation type, Wind Speed, Barometric Pressure, and Solar radiation. Hence, these two data sets were allocated for identifying Inter-relationships of weather variables that affect or are affected by Temperature and Precipitation uncovered through the correlation analysis in Section 7.

3 Precipitation Calculation

In the process of calculation of precipitation, a few methods could be employed depending on the temporal resolution of the available data.

3.1 Standard Precipitation Index (SPI)

SPI is a measure of how likely a certain amount of precipitation is to occur in a given period. It involves fitting the precipitation data to a curve that describes how often different amounts of precipitation occur, and then finding the corresponding point on a normal curve that has the same probability. The gamma curve is usually the best choice for fitting the precipitation data. The density function expression for this distribution is as follows [10]:

$$g(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} \tag{1}$$

Here, α - shape parameter, β - scale parameter, x - the amount of precipitation, and $\Gamma(\alpha) = \int_0^{\infty} y^{\alpha-1} e^{-1} dy$ is gamma function. The maximum likelihood estimates of the parameters α and β are:

$$\hat{\alpha} = \frac{1}{4A}\left(1 + \sqrt{\frac{4A}{3}}\right) \tag{2}$$

$$\hat{\beta} = \frac{\bar{x}}{\hat{\alpha}} \tag{3}$$

Here, $A = ln(\bar{x}) - \frac{\Sigma ln(x)}{n}$ and n is the sample size. By integrating formula (1), which is the distribution function of precipitation G(x), we have the following expression:

$$G(x) = \int_0^x \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} dx, (x > 0), \tag{4}$$

where $G(x)$ shows the chance of getting a precipitation amount that is equal to or lower than x.Sometimes the actual precipitation samples may have a value of 0, which means no precipitation. This means that the curve for precipitation has to be adjusted. The rectified distribution function is as follows [10]:

$$H(x) = q + (1-q)G(x) \tag{5}$$

where q is the chance of getting no precipitation at all. Based on the normal curve, the function that shows how likely different amounts of precipitation are is given by [10]:

$$\Phi(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \tag{6}$$

where $F(t)$ shows the chance of getting a value that is equal to or lower than t. The actual precipitation amount gives the related probability $H(x)$. To find the SPI value, we need to solve for t in the equation: $H(x) = (t)$. This gives the following formula :

$$SPI = t = \Phi^{-1}(H(x)) \tag{7}$$

Therefore, $H(x)$ is equivalent to a standard normal variable t that has a mean of zero and a variance of one. This means that SPI follows the standard normal distribution. The formula above indicates that computing SPI requires a sufficient number of precipitation samples, usually more than 30 years of data [11] to estimate a reliable q value and the two parameters α and β [10].

3.2 Estimating Precipitation Distribution Parameters for Stations with Short Data Series

It is hard to compute SPI without long-term observation data. Research suggests that the parameters of precipitation distribution for stations with short data series can be derived from the parameters of precipitation distribution for nearby stations with long data series as follows [10]:

1. **Nearest Neighbor Substitution Method:** This method calculates the spatial distance between the stations based on their latitude and longitude data. Then, it assigns the parameters of precipitation distribution for the short-sequence station to be the same as the parameters of the closest long-sequence station.
2. **Regional Average Method:** This method selects N long-sequence stations around the short-sequence station based on the spatial distance information and a given value of N. Then, it computes the average values of the parameters of precipitation distribution for the N long-sequence stations and uses them to calculate the SPI for the short-sequence station.
3. **Kriging Interpolation Method:** This method is a spatial auto co-variance optimal interpolation method that is widely used in spatial interpolation problems in Geosciences. It applies the ordinary Kriging interpolation method (which assumes a linear relationship between the semi-variance function and the distance) to estimate the parameters of precipitation distribution using the same data as the regional averaging method.

4 Atmospheric Temperature in the Tropopause

The sun's radiation and altitude determine the atmospheric temperature. As the altitude increases, the atmospheric temperature decreases because the sunlight has less effect on warming the Earth. The boundary layer where the temperature is balanced is called tropopause. Above the tropopause is the stratosphere, which warms up from above. The troposphere extends from the ground level to about 16 km (53,000 feet) in altitude. The stratosphere reaches up to 50 km (164,000 feet), just above the ozone layer. To calculate the atmospheric temperature, we need to know the specific location on the earth and the altitude. The average temperature in the troposphere drops by 6.5 degrees Celsius per kilometer or 3.5 degrees F per 1,000 feet of altitude. For example, at 5 km above the ground, the temperature would be $15 - (5 \times 6.5) = -17.5$ degrees Celsius. This formula is fairly accurate up to the tropopause [12].

5 Correlation Analysis Towards Identifying Interplay Between Temperature and Precipitation

Correlation analysis is important for identifying the inter-relationship between different variables. The elements that are off the diagonal reveal various insights [19, 20, 21]. The temperature and precipitation variations were observed in Figure 1 which depicts data from 1990 to 2002 in Almaty, Kazakhstan-Central Asia which experienced an extreme continental climate.

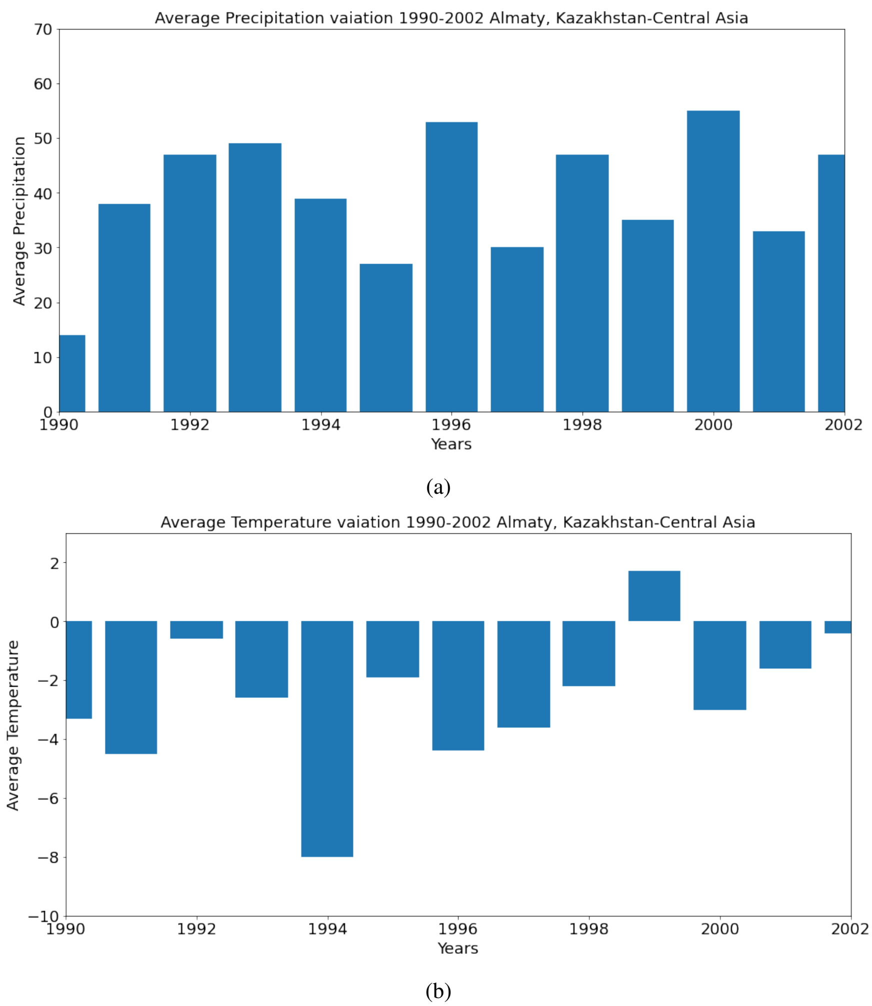


Fig. 1: Average variation of (a) Precipitation (b) Temperature from 1990-2003 Almaty, Kazakhstan.

Furthermore, we also use Dublin, Beauceville, and Alpena data sets that have a better temporal resolution of 24 hours, and we use them for carrying out correlation analysis. The variations in Figure 2 have been observed using the pattern of precipitation and temperature oscillation with the temporal resolution of 24 hours. The temperature oscillated closer to a $|sinx|$ wave yearly depending on seasonal changes. The orange line graph shows the temperature variation and the Blue line graph shows the precipitation variation.

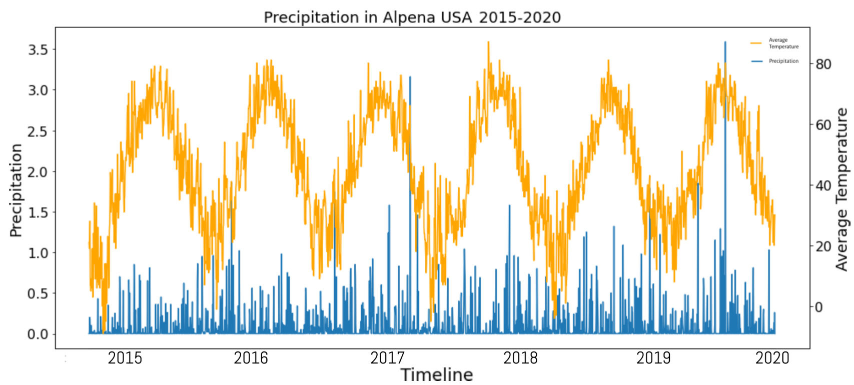


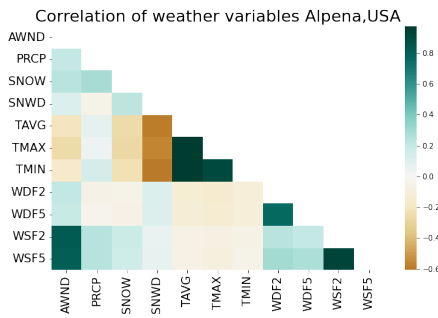
Fig. 2: Alpena - USA Temperature and Precipitation Oscillation from 2015 to 2020

Correlation analysis is significantly important in figuring out the interplay of temperature and precipitation. In statistics,

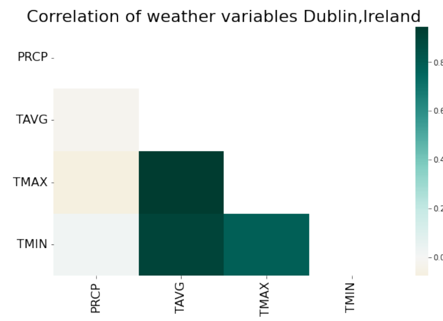
$$Correlation = \sigma = \frac{cov(X,Y)}{\sigma x \sigma y} \quad (8)$$

When one variable changes, the other variable changes by the same magnitude in the same direction, which shows a positive correlation between them [13]. When both variables change in opposite directions, it shows a negative correlation. A value between 0 and -1 indicates that the two securities have opposite movements. The correlated variables have a perfect negative correlation when σ equals -1. Correlation analyses of different weather datasets from the world are given below. Here we have selected the areas Alpena - United States, Beauceville - Canada, Dublin - Ireland, Sydney-Australia with a temporal resolution of 24 hours, and Chicago -United states with a temporal resolution of one hour [9].

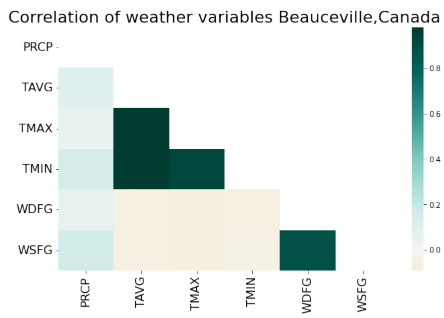
The observations of the correlation graphs (cf. Fig. 3) show a negative correlation between temperature and precipitation. We also observe that wind direction demonstrates a smaller positive correlation with precipitation, whereas wind speed shows a negative relationship. Snow is positively correlated to precipitation and negatively correlated to temperature, confirming the aforementioned discovery.



(a) Alpena USA



(b) Dublin-Ireland



(c) Beauceville-Canada

Fig. 3: Correlation Analysis of weather parameters around three distinct cities around the world.

6 Rainfall Prediction

Using multiple linear regression and long-short-term memory, this section shows how weather parameters can predict temperature and precipitation and how temperature and precipitation can predict rainfall. Since the Dublin, Alpena, and Beauceville datasets did not contain direct rainfall data, we considered precipitation as a major factor related to causing rainfall. We used the Chicago Lake Michigan lakefront dataset with the parameters of Precipitation Type, Wind Speed, Air-Temperature, Solar Radiance, Rain Intensity, Wet Bulb Temperature, Humidity, Barometric Pressure, Total rain, Interval Rain, Wind Direction, and Maximum Wind Speed; and Sydney- Australia Dataset which contains the parameters - Minimum and maximum temperatures, rainfall, evaporation, wind data, humidity, pressure, and cloud data with a temporal resolution of two hours and twenty-four hours, respectively, for the rainfall prediction.

6.1 Rainfall Prediction using Multiple Linear Regression

Regression analysis is a statistical method that estimates how variables are related when one variable causes another variable to change. Univariate regression aims to find a linear equation that relates one dependent variable and one independent variable. The dependent variable is the outcome and the independent variable is the predictor [14]. A Multiple Linear Regression analysis is performed in predicting the values of a dependent variable, Y, given a set of p explanatory variables $(x1, x2, ..., xp)$ [15]. In the first exercise, we consider temperature as the dependent variable; and average wind speed, snow depth, and precipitation are the independent variables. In the second exercise, the dependent variable is precipitation, and temperature, wind speed, and snow depth are the independent variables.

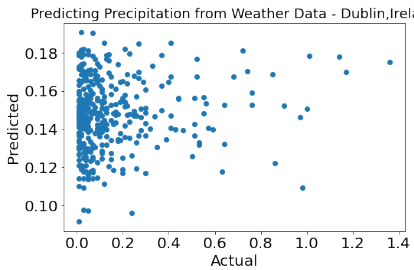
Multiple linear regression was the first method used in this study to predict temperature using other variables available in the dataset. We observe that the predicted results were significantly accurate with the actual values. This also confirms the theory of temperature's effect on the weather variables of precipitation, humidity, and pressure. When multiple linear regression was applied to predict precipitation, results were observed, as mentioned in Figure 4 were observed. Here we can see that for most of the lower precipitation values, a slightly higher value has been predicted by the model. This could be a result of multicollinearity, which is the stronger dependency of predictors with the related variables and outcome, which causes problems in estimating the regression coefficients. A rain gauge was used to measure this precipitation in inches as rainfall [9].

Therefore, we used PRCP as the settlement feature for predicting rainfall in Dublin, Alpena, and Bauceville in this study. Rainfall data has been directly employed in predicting rainfall in Sydney, Australia dataset which showed a false prediction towards the lower rainfall values and more accurate predictions for higher rainfall values as mentioned in Figure 5.

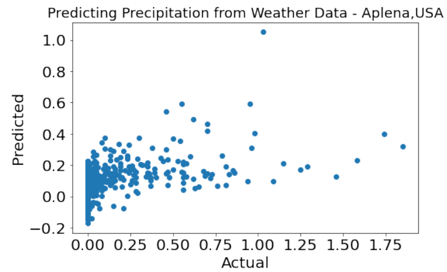
Figure 6 depicts the prediction of temperature using weather parameters such as Average Wind Speed, Snow Depth and Precipitation. The Prediction of Precipitation was conducted using Multiple Linear Regression using the same data set. The data sets were segmented into 30% test and 70% training parts. The accuracy of the temperature predictions is displayed in Table 1 which is generated from Figure 6.

Location	Alpena-USA	Chicago-USA	Beauceville-Canada	Dublin-Ireland
Coefficient of Determinations	0.98	0.99	0.99	0.95

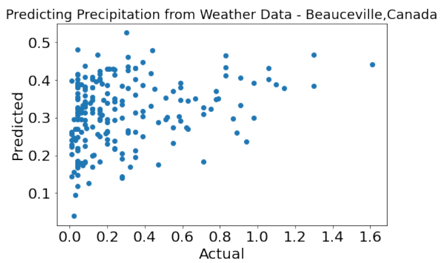
Table 1: Accuracy comparison between different data sets in predicting Temperature



(a) Dublin-Ireland



(b) Alpena - USA



(c) Beauceville - Canada

Fig. 4: Precipitation Prediction with Multiple Linear Regression

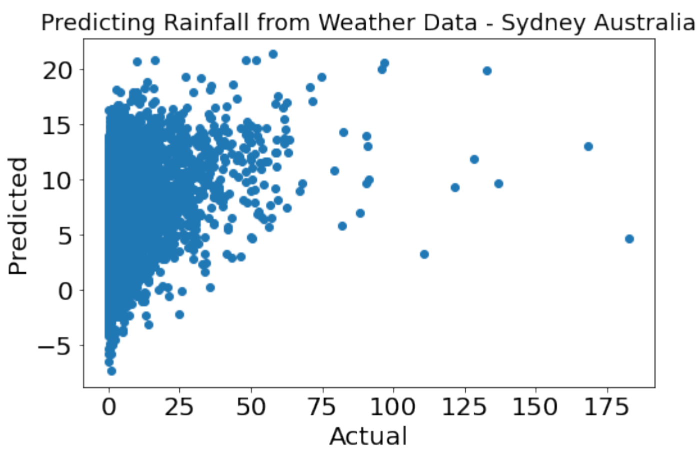
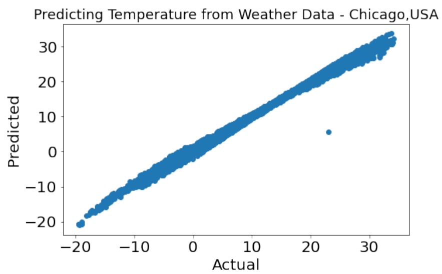


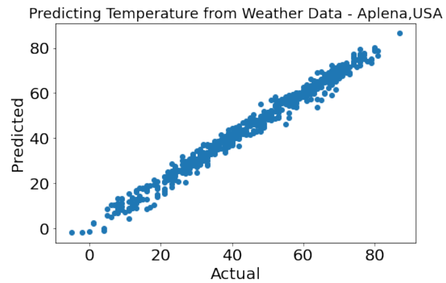
Fig. 5: Rainfall prediction for Sydney, Australia using multiple linear regression.

6.2 Rainfall Prediction using LSTM

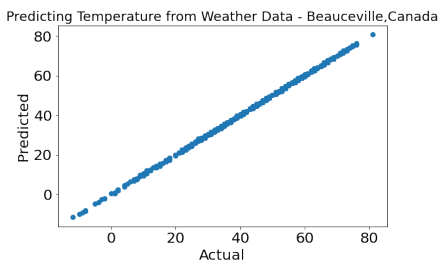
The Long-Short-Term-Memory (LSTM) model is a Recurrent Neural Network that can handle long-term dependencies better [16, 17]. All recurrent neural networks



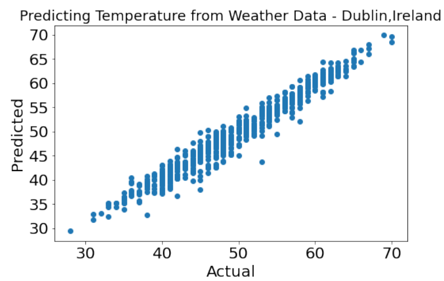
(a) Chicago - USA



(b) Alpena - USA



(c) Beauceville - Canada



(d) Dublin - Ireland

Fig. 6: Temperature prediction using multiple linear regression.

have a structure of a chain that repeats the modules [18]. LSTMs are similar to RNNs, yet contain a different structure by having four network layers instead of having a single layer, which interacts in a special way. The model has been trained with 6 epochs of batch size 32, which achieved an F1 score [22] of 0.998. The accuracy of rainfall prediction methods is usually measured by how well they detect true events and avoid false ones. [4]. Hence, the model has been developed with higher accuracy. The value and value loss achieved in each epoch are depicted in Figure 7. When comparing the two approaches for the prediction of Rainfall using Temperature and Precipitation, the accuracy was measured. In the study, multiple linear regression showed a lower accuracy whereas LSTM showed a higher accuracy. Hence we can conclude that for the used data set, LSTM was a better fit. The LTSM model was able to deal with the vanishing gradient problems [23, 24].

Prediction method	Correlation coefficient	Error %
Multiple Linear Regression	0.13	87%
LSTM	0.998	0.2%

Table 2: Accuracy comparison for different Machine Learning models for predicting rainfall in Sydney-Australia

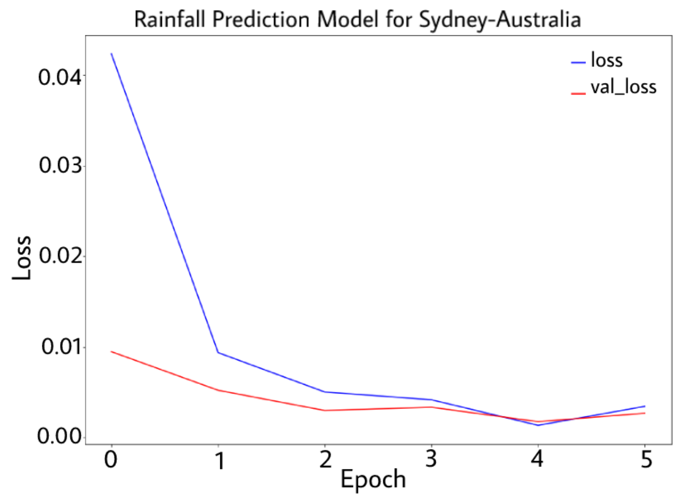


Fig. 7: Accuracy analysis of Rainfall Prediction for Sydney-Australia using LSTM.

7 Conclusion and Future Work

We have examined how temperature and precipitation affect the likelihood of rainfall in this paper. For this purpose, we have used data sets from Alpena, Chicago, Beauceville, Sydney, and Dublin. This data has been analyzed focusing on the detection of the correlation between temperature and precipitation and their interdependency with other weather parameters. These data sets have also been employed for predicting the occurrence of rainfall using different prediction models using machine learning. The accuracy of the developed techniques has been analyzed for finding the most accurate model fit for the data sets used.

In our future work, we expect to broaden with study by adopting different machine learning techniques to predict weather variables as well as rainfall and to identify the best-fitting models depending on the length of the data set and availability of data. Consequently, it will be helpful for identifying the interplay between variables toward highly accurate rainfall prediction.

Acknowledgement

This research was conducted with the financial support of Science Foundation Ireland under Grant Agreement No. 13/RC/2106_P2 at the ADAPT SFI Research Centre at University College Dublin. ADAPT, the SFI Research Centre for AI-Driven Digital Content Technology is funded by Science Foundation Ireland through the SFI Research Centres Programme.

References

1. Ren, Y.Y., Ren, G.Y., Sun, X.B., Shrestha, A.B., You, Q.L.: Observed changes in surface air temperature and precipitation in the Hindu Kush Himalayan region over the last 100-plus years. Applied Optics **8**, 148–156 (2017)
2. Fathima, T. A., Nedumpozhimana, V., Lee, Y. H., Winkler, S., Dev S.: A Chaotic Approach on Solar Irradiance Forecasting. In: Proc. Progress In Electromagnetics Research Symposium (PIERS) (2019)
3. Manandhar, S., Dev, S., Lee, Y. H., Meng, Y. S.: On the Importance of PWV in Detecting Precipitation. In: Proc. IEEE AP-S Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting (2018)
4. Manandhar, S., Dev, S., Lee, Y.H., Meng, Y.S., Winkler, S.: A data-driven approach for accurate rainfall prediction. IEEE Transactions on Geoscience and Remote Sensing **57**(11), 9323–9331 (Nov 2019). arXiv: 1907.04816
5. Adler, R.F., Huffman, G.J., Chang, A., Ferraro, R., Xie, P.P., Janowiak, J.,Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., Gruber, A., Susskind, J.,Arkin, P., Nelkin, E.: The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979 present). Journal of Hydrometeorology, vol. 4, issue 6, p. 1147 (2003).
6. Mills, F.L., Imama, E.A.: Rainfall prediction for agriculture and other resource management in the united states virgin islands. Water Resources Research Center - University of the Virgin Islands (1990)
7. Jacobs, K.L., Street, R.B.: The next generation of climate services. Climate Services **20**, 100199 (2020). https://www.sciencedirect.com/science/article/pii/S2405880720300510
8. Williams, M.W., Konovalov., V.G.: Central asia temperature and precipitation data, 1879-2003. Boulder, Colorado USA. NSIDC: National Snow and Ice Data Center https://doi.org/10.7265/N5NK3BZ8
9. Pathan, M.S., Wu, J., Nag, A., Dev, S.: A systematic analysis of meteorological parameters in predicting rainfall events p. 22
10. Zuo, D., Hou, W., Wu, H., Yan, P., Zhang, Q.: Feasibility of calculating standardized precipitation index with short-term precipitation data in China. Atmosphere **12**(5), 603 (May 2021), https://www.mdpi.com/2073-4433/12/5/603
11. Thomas, B., McKee, N., J., D., Kleist, J.: The relationship of drought frequency and duration of time scales (1993)
12. Nyveen, L.: Tutorial of how to calculate altitude & temperature (2019), https://doi.org/10.1093/acrefore/9780190228620.013.730
13. Nickolas, S.: What do correlation coefficients positive, negative, and zero mean (2021)
14. Kaya Uyanık, G., Güler, N.: A study on multiple linear regression analysis. Procedia - Social and Behavioral Sciences **106**, 234–240 (12 2013).
15. Tranmer, M., Elliot, M.: Multiple linear regression (2018)
16. Jain, M., Yadav, P., Lee, Y. H., Dev, S.: Improving Training Efficiency of LSTMs While Forecasting Precipitable Water Vapours. In: Proc. Progress In Electromagnetics Research Symposium (PIERS) (2021)
17. Pathan, M. S., Jain, M., Lee, Y. H., AlSkaif, T., Dev, S.: Efficient Forecasting of Precipitation Using LSTM. In: Proc. Progress In Electromagnetics Research Symposium (PIERS) (2021)
18. Hochreiter, S., Schmidhuber, J.: Long short term memory (1997), http://www.bioinf.jku.at/publications/older/2604.pdf
19. Manandhar, S., Dev, S., Lee, Y.H., Winkler, S., Meng, Y.S.: Systematic study of weather variables for rainfall detection. In: Proc. International Geoscience and Remote Sensing Symposium (IGARSS) (2018)
20. Dev, S., Lee, Y. H., Winkler, S.: Systematic Study of Color Spaces and Components for the segmentation of sky/cloud images. In: Proc. IEEE International Conference on Image Processing (ICIP) (2014)

21. AlSkaif, T., Dev, S., Visser, L., Hossari, M., Sark, W. van: A systematic analysis of meteorological variables for PV output power estimation. Renewable Energy (2020)
22. Sasaki, Y.: The truth of the F-measure (2007)
23. Hu, Y., Huber, A.E.G., Anumula, J., Liu, S.: Overcoming the vanishing gradient problem in plain recurrent networks. CoRR **abs/1801.06105** (2018), `http://arxiv.org/abs/1801.06105`
24. Jain, M., Manandhar, S., Lee, Y. H., Winkler, S., Dev, S.: Forecasting Precipitable Water Vapor Using LSTMs. In: Proc. IEEE AP-S Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting (2020)