**WEATHER FORECASTING USING MACHINE LEARNING**

**Table of Contents**

[Chapter 2: Background 3](#_Toc182224514)

[2.1 Introduction 3](#_Toc182224515)

[2.2 Journal Collection Process 3](#_Toc182224516)

[2.3 Critical Analysis of Literature 3](#_Toc182224517)

[2.4 Theories and Models 8](#_Toc182224518)

[2.5 Conceptual Framework 9](#_Toc182224519)

[2.6 Literature Gap 9](#_Toc182224520)

[2.7 Summary 10](#_Toc182224521)

[Chapter 3: Dataset 11](#_Toc182224522)

[Chapter 4: Ethical Issues 12](#_Toc182224523)

[Chapter 5: Methodology 13](#_Toc182224524)

[Reference List 15](#_Toc182224525)

# Chapter 2: Background

## 2.1 Introduction

Weather forecasting has been an important application of machine learning algorithms. The models are developed based on historical data of various weather parameters. The parameters could be temperature, humidity, wind speed, and other atmospheric parameters. This way the possible long-term weather prediction can be done using machine learning algorithms. This background chapter will critically analyse the existing papers on weather prediction and identify the possible literature gap. The relevant theories and models will also be discussed.

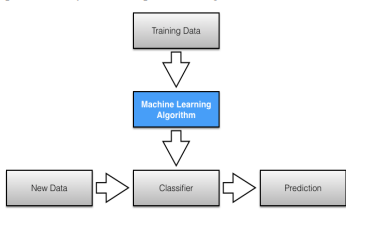
## 2.2 Journal Collection Process

The journal collection inclusion and exclusion criteria are as follows:

1. It was ensured that the research papers have been published in recent years. This is the reason why the research papers published in 2019 have been chosen for the background analysis.
2. The research papers about weather forecasting were only selected to ensure that the background analysis would be within the scope of the research.
3. The research papers written in the English language were selected as part of this research procedure.

## 2.3 Critical Analysis of Literature

As per Singh, Chaturvedi and Akhter, (2019), the constant climate change has ensured that the old weather prediction model gets obsolete. The main aim of the paper was to develop a model for weather prediction that can be used in remote areas. The authors have developed the model using “Random Forest Classification” algorithms. The result was a portable and low-cost weather prediction solution for remote areas.



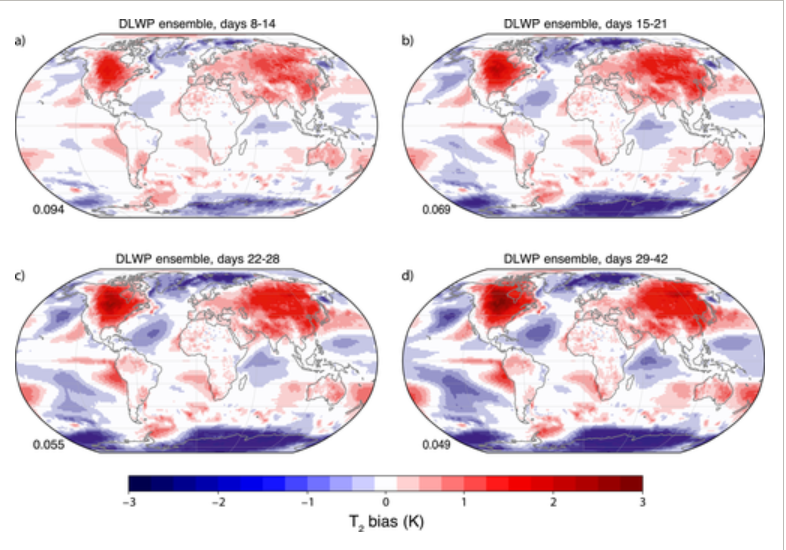
**Figure 2.1: Proposed Weather Forecasting Model**

(Source: Singh, Chaturvedi and Akhter, 2019)

The figure shows the model's overall flowchart or system architecture. The model would first classify and then make the prediction. The developed model has an accuracy of more than 87% in terms of its prediction of whether there would be rain or not. Despite the high accuracy, the limitation of the model could be that it's only evaluated based on whether it would rain or not. The evaluation based on other weather incidents had not been done. On the other hand, Jakaria, Hossain and Rahman, (2020), the weather parameters and data used in weather forecasting are unstable. The authors have developed a weather forecasting model that can predict short weather based on data from multiple weather stations. The author has collected data from various weather stations in Nashville. The authors have compared results from “Random Forest Classifier” with SVM, “Multi-layer Perceptron”, “Extra-Tree Regression”, and “Ridge Regression”. The random first classifier has shown the best result. The main strength of the paper was its usage of multiple ML algorithms for weather prediction, however, the authors have not developed a full-fledged application which is a major limitation of the model.

Bochenek and Ustrnul, (2022), conducted a systematic literature review on various weather prediction papers. The aim was to identify the most common factors and methods used for weather prediction. The result showed that radiation, pressure, temperature, precipitation, and wind are the most common meteorological fields examined for predicting weather. SVM, random forest, XGBoost, “Artificial Neural Networks” etc., have been the most popular algorithms that were used. The paper has summarised and identified the important aspects of weather parameters and algorithms used for weather prediction. However, it has not developed its own model which is a major limitation.

As per Wang *et al.* (2019), the inappropriate settings of the initial states can lead to unsatisfactory results in weather prediction. This is why the authors have proposed a data-driven weather prediction model based on machinimas of information fusion. The authors have used deep learning algorithms based on the “novel negative log-likelihood error” problem. It is able to forecast for both uncertainty quantification and single-value prediction. The accuracy of the model significantly increased compared to traditional numerical models for weather prediction. However, accuracy is still low in this regard. On the other hand, Weyn *et al.* (2021), have developed an “ensemble prediction system” based on DL algorithms. The CNN algorithm was used for six-week forecasts.



**Figure 2.2: Weather Forecasting of Cyclone**

(Source: Wang *et al.,* 2019)

This figure shows the cloud and weather prediction for Cyclone Irma using the proposed model. The proposed model can forecast 320 times within 3 minutes and this shows the effectiveness of the model. The main effectiveness of the mode is that it has been evaluated based on real-world incidents like Irma. However, it is not effective for long-term weather prediction for the future.

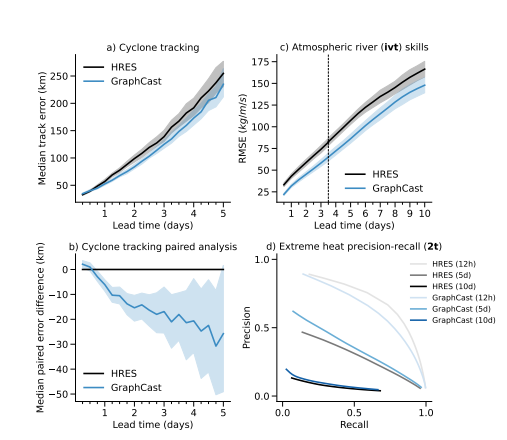
Cho *et al.* (2020), have developed a numerical model for weather prediction for extreme air temperature events in urban areas. The prediction of minimum and maximum temperature is the most basic and essential part of weather forecasting. However, coarse grading and the absence of proper parameterisation have affected the quality of prediction. The authors have used SVMM, ANN, Random Forest, and “multi-model ensemble” algorithms for predicting the minimum and maximum temperature of the next day in South Korea. The model has an R2 value of 0.69 which is more than 0.50. This shows the model has some errors despite the low RMSE value.

As per Hewage *et al.* (2021), the model of “numerical weather prediction” needs significant power for solving mathematical equations. The authors have proposed a novel “lightweight data-driven weather forecasting model” for weather forecasting. They have used the LSTM algorithm and TCN or “temporal convolutional networks” for model development. Further other classification and regression models have also been used. The forecasting of the weather is collected by the time series data in this regard. The result showed that the model has a higher accuracy compared to general deep learning models. It can forecast weather for up to 12 hrs.

As per Chattopadhyay, Nabizadeh and Hassanzadeh, (2020), the ever-growing resources and time consumption have been the main problems associated with weather prediction based on numerical values. The authors have proposed an analog forecasting model based on data. The proposed model has used CapsNet and “capsule neural networks” to build a “novel deep learning pattern-recognition technique”. The data was collected from the Earth system model and it was used to train the weather model. The accuracy of the trained model is 45% which is not satisfactory. However, the usage of CapNets has increased the model accuracy up to 80%. CapsNet has outperformed both CNN and normal logistic regression models in this regard. However, the model accuracy is still not close to 100% which is a major model limitation.

As per Chantry *et al.* (2021), ML algorithms can be valuable as an accelerator for the scheme of parameterisation. The authors have developed a parameterisation scheme using ML algorithms. The authors have used deep learning algorithms like ANN in this regard. The result showed that for the medium-range prediction, the model is highly accurate. However, for short-time and long-time prediction, the model is not that accurate for weather prediction.

Lam *et al.* (2023), that global weather forecasting is tough for medium-time forecasting due to various social and economic domains. The traditional methods need significant computing resources for prediction. The authors have developed an ML-based model called Graphcast. It can predict more than 100 weather variables at the same time. The accuracy of the model is 90% which is significantly higher than other systems of operational deterministic. It can predict severe events accurately including extreme temperatures, cyclones, atmospheric rivers and so on. This way it has been effective to develop complex models for dynamic systems.



**Figure 3: Weather Prediction using GraphCast**

(Source: Lam *et al.,* 2023)

The figure shows possible prediction graphs of Graphcats based on cyclone tracking, atmospheric rivers, extreme heat and precision levels.

As per Grönquist *et al.* (2021), the quantification of forecasting of weather is very difficult due to the extreme weather events. Ensemble prediction systems. Can be used to predict extreme weather events. However, these are costly with high computational resources. The authors have proposed a mix of traditional models and ensemble prediction models to assess the non-linear relationship between different weather parameters. The authors have used global data for their prediction methods. The authors have used case studies to prove improved weather forecasting for extreme events. This way the overall cost and resources of the ensemble system can be reduced through the mixed method. Han *et al.* (2022), have developed a wind speed prediction model as part of the quantitative weather forecasting model. The prediction of wind speed is tough due to the relationship between different meteorological parameters. The authors have used weather forecasting and research to develop a hybrid model for wind speed prediction. The authors have used a hybrid model of the “multivariate data decomposition method” and the DL model for the development of the prediction model. The deep learning algorithms were developed using CNN and bidirectional LSTM. It was found that the model proposed by the authors has outperformed other similar models including an MAE value of 0.1042. The accuracy has also been increased significantly in this regard.

Rasp *et al.* (2020), have developed a data-based forecasting model for weather prediction. The authors have tried to predict global weather based on their data-driven approach. The authors have collected datasets from the ERA5 archive. They have used simple linear regression, physical models, and deep learning algorithms for model development. The method has been able to predict weather for 3-5 days. The developed model has significantly increased the overall quality prediction quality. The lack of a detailed description of evaluation metrics could be the main limitation of this study.

## 2.4 Theories and Models

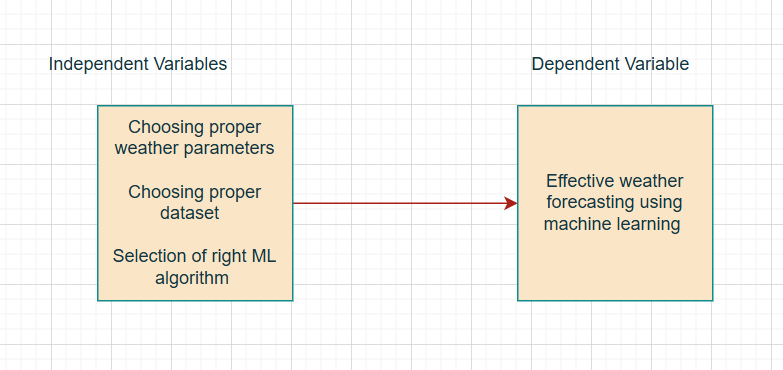
***Technology Acceptance Model (TAM)***

TAM can be defined as the model that describes how people accept a new technology. It was established by Fred Davies to show why people can accept or reject new technology. The use of modern machine learning, deep learning, and time series algorithms has improved the weather prediction quality significantly. People can see weather forecasting from their own mobile or computers. This is why weather forecasting using modern technologies can be explained using TAM.

1. ***Perceived usefulness:*** A user will use a new technology if they believe the new technology will improve job performance (Zaineldeen *et al.,* 2020). Modern techniques like Time series models, ML and DL algorithms have improved the weather forecasting performance significantly well. The audiences have easily accessed the weather prediction application from their preferred devices.
2. ***Ease of use:*** This means the user will accept a new technology if it frees their effort. The modern technologies of weather forecasting have reduced the effort of users to accept weather forecasting. The models have improved the quality of the forecasting and reduced the possible effort of researchers to collect data as well.

This shows the Ml algorithms have significantly been effective for forecasting the future weather and collecting weather data in this regard.

## 2.5 Conceptual Framework



**Figure 2.4: Conceptual Framework**

(Source: Created by the Authors)

This is the conceptual framework that this research procedure would like to follow. Here, the “dependent variable” is “Effective weather forecasting using machine learning”. Now, the “independent variables” are choosing proper weather parameters, choosing the proper dataset selecting of right ML algorithm.

## 2.6 Literature Gap

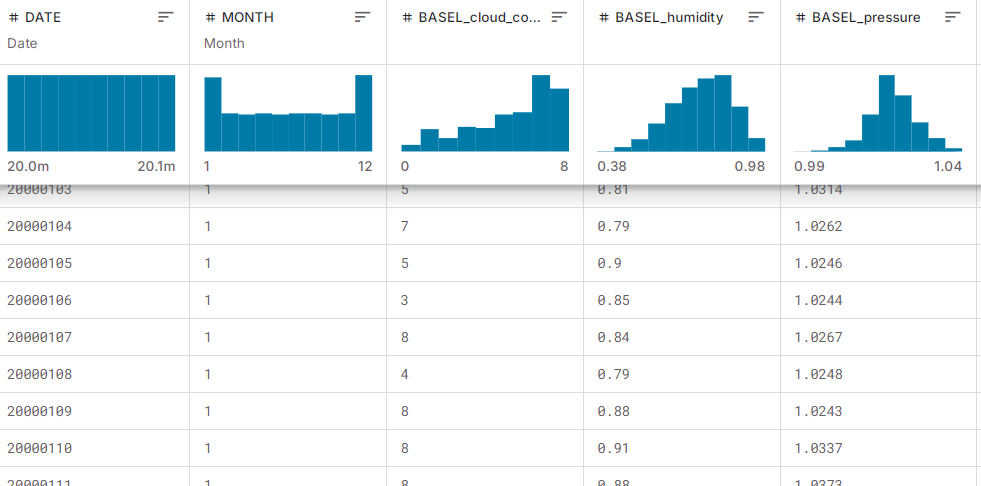
The various research papers have discussed the usage of various algorithms and various weather parameters for weather forecasting. However, there has been a lack of specificity about which weather parameters researchers need to consider to focus their models on. There are specific weather and atmospheric parameters like cloud formation, wind speed, temperature, pressure, humidity etc. This research procedure has the purpose of explaining in depth the specific weather parameters that will be used for predicting the weather in this regard. This way the specific algorithms will be selected for developing the model. This way the classification and prediction procedure will be easier.

## 2.7 Summary

To summarise, the background section has provided the required information to assess the existing papers about weather prediction. Random Forest has been the most popular algorithm that has been used for forecasting weather followed by SVM and other deep learning algorithms. The data-driven approach for numerical prediction of weather has been a trend in recent years. The technological acceptance model has shown why both researchers and users are accepting modern technology for weather forecasting. Based on the analysis it was found that the lack of explanation of the weather parameters has been the primary limitation of the existing study. This will be mitigated in this study.

# Chapter 3: Dataset

This research procedure will use a combination of two datasets. The first dataset was collected from the Kaggle website as it is a popular data collection website. The dataset link is as follows: “<https://www.kaggle.com/datasets/thedevastator/weather-prediction/data?select=weather_prediction_dataset.csv>”. The other dataset that was collected was from Cambridge University’s website. The link is “<https://www.cl.cam.ac.uk/weather/>”.



**Figure 3.1: Weather dataset**

(Source: Kaggle, 2024)

This is the collected data from the Kaggle website. The columns consist of humidity, maximum and minimum temperature, global radiation, pressure etc. Originally the data was collected from 18 different regions of Europe. It was developed for training the model in this regard.

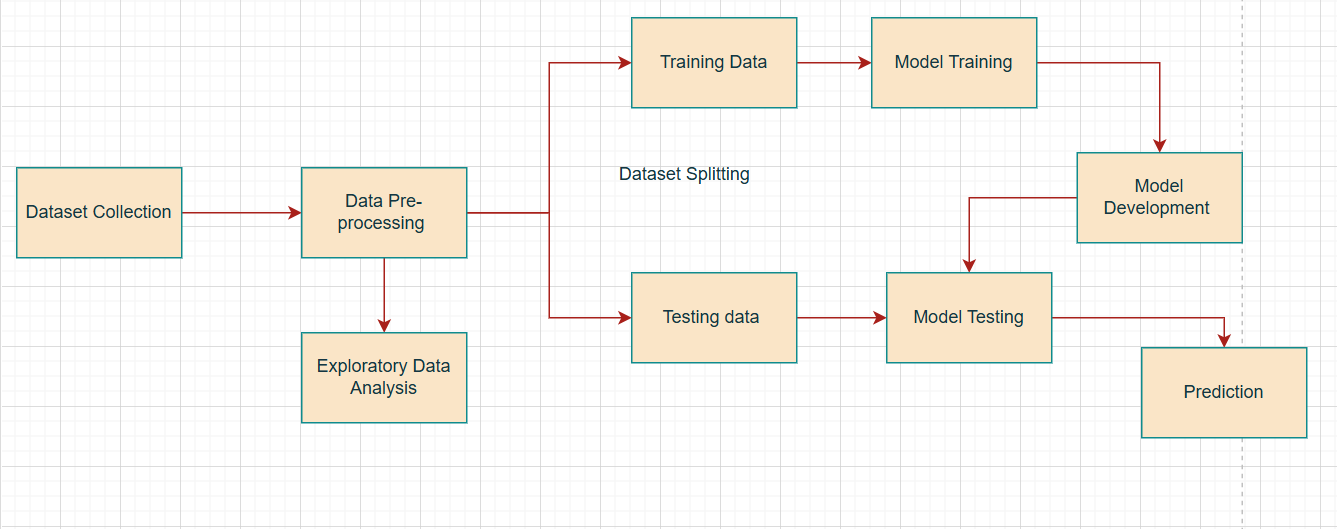
# Chapter 4: Ethical Issues

The dataset does not have any personal data of people and it is publically available to use. This is the reason why it does not have any privacy issues for data breaches. There is no chance of breaching the GDPR principle because it does not have any data privacy issues.

The dataset has indeed used several journal papers for background study. The target will be to appropriately reference all those secondary sources of data so that the original data owner gets the credit (Hummel, Braun and Dabrock, 2021). One of the major considerations of the research procedure will be to honestly present all the data. The findings from the software will be honestly presented with screenshots of the software in this regard.

# Chapter 5: Methodology

***System Architecture***



**Figure 5.1: System Architecture**

(Source: Self-created using Draw.io)

This is the overall system architecture that was used in this research procedure.

1. **Dataset Collection:** The dataset was collected in this stage. The details of the weather data collection have already been explained previously.
2. **Data Pre-processing:** The dataset was then prepared based on various procedures. This includes trying to find the null values or missing values and so on. Further, the main weather parameters were identified and features were extracted during this stage.
3. **Exploratory Data Analysis:** The overall dataset was visualised to assess the nuanced understanding of it.
4. **Dataset Splitting:** The dataset was divided for training and testing. Most of the data was used for training and the others were used for testing the model.
5. **Model Training:** the model was then trained using the training data. It was ensured that the model was not over fitted.
6. **Model Development:** The selected machine learning model was then developed based on the collected data.
7. **Model Testing:** The testing data was used during this stage to test the model. The model evaluation techniques like classification reports were used for testing the model in this stage.
8. **Prediction:** Finally the developed model predicts the weather and fulfils the main objective of the research.

# Reference List

Bochenek, B. and Ustrnul, Z., 2022. Machine learning in weather prediction and climate analyses—applications and perspectives. Atmosphere, 13(2), p.180. [https://www.mdpi.com/2073-4433/13/2/180](https://www.mdpi.com/2073-4433/13/2/180%20)

Hummel, P., Braun, M. and Dabrock, P., 2021. Own data? Ethical reflections on data ownership.

Philosophy & Technology, 34(3), pp.545-572. <https://link.springer.com/content/pdf/10.1007/s13347-020-00404-9.pdf>

Chantry, M., Hatfield, S., Dueben, P., Polichtchouk, I. and Palmer, T., 2021. Machine learning emulation of gravity wave drag in numerical weather forecasting. Journal of Advances in Modeling Earth Systems, 13(7), p.e2021MS002477. [https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2021MS002477](https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2021MS002477%20)

Chattopadhyay, A., Nabizadeh, E. and Hassanzadeh, P., 2020. Analog forecasting of extreme‐causing weather patterns using deep learning. Journal of Advances in Modeling Earth Systems, 12(2), p.e2019MS001958. [https://agupubs.onlinelibrary.wiley.com/doi/pdfdirect/10.1029/2019MS001958](https://agupubs.onlinelibrary.wiley.com/doi/pdfdirect/10.1029/2019MS001958%20)

Cho, D., Yoo, C., Im, J. and Cha, D.H., 2020. Comparative assessment of various machine learning‐based bias correction methods for numerical weather prediction model forecasts of extreme air temperatures in urban areas. Earth and Space Science, 7(4), p.e2019EA000740. [https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2019EA000740](https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2019EA000740%20)

Grönquist, P., Yao, C., Ben-Nun, T., Dryden, N., Dueben, P., Li, S. and Hoefler, T., 2021. Deep learning for post-processing ensemble weather forecasts. Philosophical Transactions of the Royal Society A, 379(2194), p.20200092. [https://royalsocietypublishing.org/doi/full/10.1098/rsta.2020.0092](https://royalsocietypublishing.org/doi/full/10.1098/rsta.2020.0092%20)

Han, Y., Mi, L., Shen, L., Cai, C.S., Liu, Y., Li, K. and Xu, G., 2022. A short-term wind speed prediction method utilizing novel hybrid deep learning algorithms to correct numerical weather forecasting. Applied Energy, 312, p.118777. <https://www.researchgate.net/profile/Guoji-Xu-2/publication/358632567_A_short-term_wind_speed_prediction_method_utilizing_novel_hybrid_deep_learning_algorithms_to_correct_numerical_weather_forecasting/links/620c6bf2afa8884cabe7a14b/A-short-term-wind-speed-prediction-method-utilizing-novel-hybrid-deep-learning-algorithms-to-correct-numerical-weather-forecasting.pdf>

Hewage, P., Trovati, M., Pereira, E. and Behera, A., 2021. Deep learning-based effective fine-grained weather forecasting model. Pattern Analysis and Applications, 24(1), pp.343-366. [https://link.springer.com/content/pdf/10.1007/s10044-020-00898-1.pdf](https://link.springer.com/content/pdf/10.1007/s10044-020-00898-1.pdf%20)

Jakaria, A.H.M., Hossain, M.M. and Rahman, M.A., 2020. Smart weather forecasting using machine learning: a case study in tennessee. arXiv preprint arXiv:2008.10789. [https://arxiv.org/pdf/2008.10789](https://arxiv.org/pdf/2008.10789%20)

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., Ravuri, S., Ewalds, T., Eaton-Rosen, Z., Hu, W. and Merose, A., 2023. Learning skillful medium-range global weather forecasting. Science, 382(6677), pp.1416-1421. [https://arxiv.org/pdf/2212.12794](https://arxiv.org/pdf/2212.12794%20)

Rasp, S., Dueben, P.D., Scher, S., Weyn, J.A., Mouatadid, S. and Thuerey, N., 2020. WeatherBench: a benchmark data set for data‐driven weather forecasting. Journal of Advances in Modeling Earth Systems, 12(11), p.e2020MS002203.[https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2020ms002203](https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2020ms002203%20%20)

Singh, N., Chaturvedi, S. and Akhter, S., 2019, March. Weather forecasting using machine learning algorithm. In 2019 International Conference on Signal Processing and Communication (ICSC) (pp. 171-174). IEEE. [https://dcs.datapro.in/wp-content/uploads/2023/11/36-Weather-Forecasting-Using-Machine-Learning.pdf](https://dcs.datapro.in/wp-content/uploads/2023/11/36-Weather-Forecasting-Using-Machine-Learning.pdf%20%20)

Wang, B., Lu, J., Yan, Z., Luo, H., Li, T., Zheng, Y. and Zhang, G., 2019, July. Deep uncertainty quantification: A machine learning approach for weather forecasting. In Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining (pp. 2087-2095). [https://arxiv.org/pdf/1812.09467](https://arxiv.org/pdf/1812.09467%20)

Weyn, J.A., Durran, D.R., Caruana, R. and Cresswell‐Clay, N., 2021. Sub‐seasonal forecasting with a large ensemble of deep‐learning weather prediction models. Journal of Advances in Modeling Earth Systems, 13(7), p.e2021MS002502. [https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2021MS002502](https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2021MS002502%20)

Zaineldeen, S., Hongbo, L., Koffi, A.L. and Hassan, B.M.A., 2020. Technology acceptance model’concepts, contribution, limitation, and adoption in education. Universal Journal of Educational Research, 8(11), pp.5061-5071. <https://www.researchgate.net/profile/Samar-Zaineldeen/publication/345004753_Technology_Acceptance_Model'_Concepts_Contribution_Limitation_and_Adoption_in_Education/links/5f9c0445299bf1b53e514fa8/Technology-Acceptance-Model-Concepts-Contribution-Limitation-and-Adoption-in-Education.pdf>

Cl.cam.ac.uk, (2024), Accessed from: https://www.cl.cam.ac.uk/weather/ [Accessed on: 09/10/2024]

Kaggle, (2024), Accessed from: https://www.kaggle.com/datasets/thedevastator/weather-prediction/data?select=weather\_prediction\_dataset.csv [Accessed on: 09/10/2024]