



Assessing the Impact of Atmospheric CO₂ Concentrations on Rainfall Patterns

R.A. Wijesinghe
(Reg. No.: MS23005112)

A THESIS
SUBMITTED TO
SRI LANKA INSTITUTE OF INFORMATION TECHNOLOGY
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE IN INFORMATION TECHNOLOGY

December 2024

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Ms. Suranjini Silva

Approved for MSc. Research Project:

MSc. Programme Co-ordinator, SLIIT

Approved for MSc:

Head of Graduate Studies, FoC, SLIIT

DECLARATION

This is to certify that the work is entirely my own and not of any other person, unless explicitly acknowledged (including citation of published and unpublished sources). The work has not previously been submitted in any form to the Sri Lanka Institute of Information Technology or to any other institution for assessment for any other purpose.

Sign: 

Date: 10/11/2024

ABSTRACT

Assessing the Impact of Atmospheric CO₂ Concentrations on Rainfall Patterns

R.A.Wijesinghe

MSc. In Information Technology

Supervisor: Ms. Suranjini Silva

December 2024

This research aims to assess the impact of atmospheric CO₂ concentrations on rainfall patterns, focusing on the relationship between key environmental parameters such as temperature, humidity, wind speed, wind direction, atmospheric pressure, and rainfall. Data were collected over 17 months, including CO₂ data sourced from the National Building Research Organization (NBRO) in Colombo and additional CO₂ measurements captured via MG811 CO₂ sensor. Environmental data such as temperature, humidity, wind speed, wind direction, and pressure were obtained from the Sri Lanka Meteorological Department, ensuring a comprehensive dataset for analysis. Machine learning algorithms, including Random Forest, XGBoost, and LSTM, were employed to develop predictive models for rainfall based on the collected data. The performance of these models was evaluated using various metrics like Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared values. Results indicate that incorporating CO₂ data improves model performance, particularly with the Random Forest model, which demonstrated the lowest error rates and highest predictive reliability when CO₂ was included as a feature. The findings underscore the importance of considering atmospheric CO₂ in climate modeling, revealing that CO₂ levels may have a more complex and region-specific influence on rainfall than previously recognized. This enhanced forecasting approach has significant implications for various sectors in Sri Lanka, including agriculture, aviation, and fisheries, where accurate rainfall predictions are critical for planning and resource management. The study's outcomes are especially relevant for policymakers and environmental stakeholders, as they highlight the potential for data-driven strategies to mitigate climate impacts and promote sustainable development practices. Additionally, the research contributes to the academic discourse on climate dynamics, offering valuable insights for future studies and serving as a foundation for educational initiatives in environmental science and meteorology.

ACKNOWLEDGEMENT

I would like to express my deepest gratitude to my supervisor, Ms. Suranjini Silva, for her invaluable advice, insightful feedback, and constant encouragement. Her guidance and support throughout every stage of this project were crucial to both the completion of this final report and my understanding of the problem domain. I am also deeply thankful to my family for their unwavering encouragement and the knowledge they shared, which helped me reach my goals. Finally, I extend my heartfelt thanks to all those who provided their assistance throughout this project. Although I cannot mention each name individually, their contributions are truly appreciated and gratefully acknowledged.

TABLE OF CONTENTS

DECLARATION	ii
ABSTRACT.....	iii
ACKNOWLEDGEMENT.....	iv
TABLE OF CONTENTS.....	v
List of Figures	viii
List of Tables.....	ix
Chapter 1 Introduction	1
1.1 Overview	1
1.2 Background	2
1.3 Research Problem.....	3
1.4 Objective	4
1.5 Significance of the Study	5
1.6 Proposed Solution	6
Chapter 2 Literature Review.....	7
2.1 Review of Existing Literature	7
2.1.1 Climate Change and CO2 Levels.....	7
2.1.2 CO2 and Rainfall.....	9
2.1.3 Temperature and Rainfall.....	11
2.1.4 Humidity and Rainfall.....	12
2.1.5 Wind speed and Rainfall	14
2.1.6 Wind direction and Rainfall	16
2.1.7 Pressure and Rainfall.....	18
2.1.8 Measure Atmospheric CO2 Concentration	20
2.1.9 Machine Learning for Weather Research.....	21
2.1.10 Integrating IoT and Machine Learning	23
2.2 Gap Identification.....	24

2.3 Conceptual Framework	25
Chapter 3 Research Objectives and Questions	27
3.1 Research Objectives	27
3.2 Research Questions	28
Chapter 4 Research Design and Methodology	30
4.1 Research Design.....	30
4.2 Sampling Strategy	32
4.3 Data Collection Methods.....	33
4.4 Data Analysis Techniques	35
4.4.1 Import the Dataset	36
4.4.2 Convert Date and Time to Standardized datetime Format	36
4.4.1 Set the Index Column.....	37
4.4.2 Handle the Missing Values	37
4.4.3 Handle the Outliers.....	38
4.4.4 Correlation among the Variables.....	38
4.4.5 Relationship between CO2 Concentrations and Rainfall	39
4.4.6 Identify the Seasonal Patterns	40
4.4.7 Data Augmentation.....	42
4.4.8 Data Preparation and Feature Selection	42
4.4.9 Random Forest Modeling.....	43
4.4.10 XGBoost Modeling	43
4.4.11 Long Short-Term Memory Modeling	44
4.4.12 Performance Evaluation	45
4.4.13 Feature Importance.....	46
4.4.14 Shapley Additive Explanations	47
4.4.15 Hyperparameter Tuning	47
4.4.16 Forecast the Values	48

4.4.17 Evaluate the Predictive Performance	49
4.5 Significance and Expected Contributions	50
4.5.1 Significance of the Study	50
4.5.2 Expected Contributions	51
Chapter 5 Results	52
5.1 Data Summary.....	52
5.2 Handle the Outliers.....	58
5.3 Correlation among the Variables.....	61
5.4 Relationship between CO2 Concentrations and Rainfall	63
5.5 Identify the Seasonal Patterns	63
5.6 Performance Evaluation	65
5.7 Feature Importance.....	67
5.8 Shapley Additive Explanations	69
5.9 Hyperparameter Tuning	70
5.10 Performance Evaluation for the Optimized Model	71
5.11 Predict the Values	72
5.12 Evaluate the Predictive Performance	73
Chapter 6 Discussion	75
Chapter 7 Conclusion.....	80
7.1 Limitations	82
7.2 Future Research Directions	83
References	85
Appendix	94
Appendix 1: Machine Learning Outputs.....	94
Appendix 2: Arduino Code	109
Appendix 3: Domain Validation	114

List of Figures

Figure 2.1 Conceptual Framework.....	25
Figure 4.1 Research Design	30
Figure 4.2 IoT Device to Collect CO2 Data.....	34
Figure 4.3 Firebase Database	34
Figure 4.4 Seasonal Identification Function	40
Figure 5.1 Distribution of Rainfall Values.....	52
Figure 5.2 Distribution of CO2 Values	53
Figure 5.3 Distribution of Temperature Values	54
Figure 5.4 Distribution of Humidity Values	55
Figure 5.5 Distribution of Wind Speed Values	56
Figure 5.6 Distribution of Wind Direction Values.....	57
Figure 5.7 Distribution of Pressure Values	58
Figure 5.8 Distribution of Humidity (After Outlier Handling)	59
Figure 5.9 Distribution of Wind Speed (After Outlier Handling).....	59
Figure 5.10 Distribution of Pressure (After Outlier Handling)	60
Figure 5.11 Correlation Heatmap.....	61
Figure 5.12 Scatter Plot for each Feature	62
Figure 5.13 Average Rainfall by Season.....	63
Figure 5.14 Random Forest Feature Importance Plot	68
Figure 5.15 SHAP Summary Plot	69
Figure 5.16 Force Plot.....	70
Figure 5.17 Actual vs Predicted Rainfall Values (Model with CO2)	72
Figure 5.18 Actual vs Predicted Rainfall Values (Model without CO2)	73

List of Tables

Table 5.1 Correlation between CO ₂ and Rainfall	63
Table 5.2 Seasonal Correlational Values	64
Table 5.3 Mean Absolute Error Values.....	65
Table 5.4 Root Mean Squared Error Values	65
Table 5.5 R ² values.....	66
Table 5.6 Performance Table for the Models.....	67
Table 5.7 Optimized Model's Performance Evaluation with CO ₂	71
Table 5.8 Optimized Model's Performance Evaluation without CO ₂	71
Table 5.9 Predictive Performance Evaluation Values with CO ₂	73
Table 5.10 Predictive Performance Evaluation Values without CO ₂	74