2025 International Conference on Artificial Intelligence, Computer, Data Sciences and Applications (ACDSA) | 979-8-3315-3562-9/25/$31.00 ©2025 IEEE | DOI: 10.1109/ACDSA65407.2025.11166651

Exploring Temporal Machine Learning Approaches for Predicting Methane Discharge in Atmospheric Studies

Sara Zebardast
Department of Information Systems and Technologies
Bilkent University
Ankara, Türkiye
sara.zebardast@ug.bilkent.edu.tr

Nima Kamali Lassem
Ecole Polytechnique de Bruxelles
Université libre de Bruxelles
Brussels, Belgium
nima.kamali.lassem@ulb.be

Seyid Amjad Ali
Department of Information Systems and Technologies
Bilkent University
Ankara, Türkiye
syedali@bilkent.edu.tr

***Abstract*—Methane levels in the atmosphere are rising, posing a serious challenge in the fight against climate change. Because methane is a powerful greenhouse gas, being able to accurately forecast its concentration is essential for making smart decisions about climate policy. Traditional ways of monitoring methane rely on direct measurements, which can be expensive and are often limited to certain areas. This study aims to explore how artificial intelligence algorithms (machine learning and deep learning) can help fill the gap. Using time series data from Alberta's methane warehouse, we tested and compared four different models: Random Forest, Extreme Gradient Boosting, Multilayer Perceptron and Temporal Convolutional Networks. We assessed how well each model predicted methane levels over time and how efficiently they handled the data. The findings offer insights into building more affordable and scalable systems for tracking and managing methane emissions, helping support better climate strategies.**

Keywords—Methane Prediction, Time Series Data, Machine Learning, Random Forest, XGBoost, Multilayer Perceptron, Temporal Convolutional Networks.

I. INTRODUCTION

The alleviation of climate change faces a critical challenge in controlling methane emissions, which have reached unprecedented levels. As documented by the Climate and Clean Air Coalition, the warming impact of methane is 86 times stronger than that of CO_2. During the last 20 years, more than 60% of methane emissions are a direct result of human activities [1]. These activities include livestock farming—particularly from digestive processes in cattle—rice paddies, fossil fuel operations like oil and gas extraction, coal mining, the decomposition of waste in landfills, and the treatment of wastewater. Other contributors include practices like biomass burning and the use of certain biofuels [2]. The urgency of this issue is highlighted by 2021's record-breaking annual increase in methane emissions, the largest since global monitoring began four decades ago. In 2021, atmospheric methane levels increased by 17 parts per billion (ppb), the largest annual increase since monitoring began in 1983 [3].

Recent research has demonstrated the far-reaching implications of methane emissions. Van Amstel's comprehensive review identifies methane as the second most important greenhouse gas, with atmospheric concentrations 2.5 times higher than preindustrial levels [4]. Beyond its direct climate impact, studies have shown methane's role in forming tropospheric ozone, which contributes to regional air pollution and poses significant health risks [5]. Statistical projections further underscore this concern, with a recent analysis by Güleryüz predicting a further 6.9% increase in methane emissions by 2026 using Double Exponential Smoothing methods, which outperformed both ARIMA and Grey Model approaches [6]. This dual impact on both climate and air quality makes accurate methane prediction crucial for environmental management.

While traditional monitoring approaches rely on direct measurements, the scientific community has increasingly turned to machine-learning techniques for methane prediction. Luo et al., [7] demonstrated the effectiveness of this approach in Alberta, showing that Long Short-Term Memory (LSTM) networks significantly outperformed conventional methods, achieving 35.38% better MAPE and 48.02% better RMSE compared to ARIMA models. Similarly, research in Türkiye has shown promising results using various machine learning approaches, with logistic regression achieving 94. 9% precision in the prediction of methane [8]. However, there remains a significant gap in understanding which machine learning and deep learning approaches are most effective for the prediction of methane concentration in different contexts and data types. This study has attempted to fill this gap by conducting a comprehensive comparison of four distinct models: Random Forest, Extreme Gradient Boosting (XGBoost), Multilayer Perceptron (MLP) and Temporal Convolutional Networks (TCNs). We evaluate these models' predictive accuracy and computational efficiency using Alberta's methane warehouse time series data [9].

II. DESCRIPTIVE STATISTICS

A. Data Collection and Preprocessing

The dataset used for training the machine learning models was obtained from the Alberta Air Data Warehouse, which provides comprehensive air quality monitoring data throughout the province. We obtained hourly time series data between

2015-2024 from two monitoring stations: Calgary Central and Red Deer-Riverside. The raw data underwent multiple preprocessing steps to ensure quality and consistency. Missing values were filled using forward filling—a technique that replaces the most recent missing value with the last observed value to fill in the gaps to preserve temporal order in a time series data [10].

B. Maintaining the Integrity of the Specifications

The dataset utilized in this study contains a total of 140,256 recorded measurements of methane concentration, which are evenly divided between the two distinct atmospheric monitoring stations. These measurements represent continuous and systematic observations collected over time to ensure balanced representation from both locations. A detailed overview of the dataset's main descriptive statistical characteristics, including the fundamental measures such as mean, standard deviation, minimum and maximum values, is presented in Table I to provide insight into the nature of the data.

1) *Basic Statistics:* The overall mean for methane concentration in the dataset is 2.075 ppm, with a standard deviation of 0.220 ppm, indicating a moderate variation in methane levels. The median and mode are both 2.000 ppm. Observed concentrations range between a minimum of 1.400 ppm to a maximum of 6.323 ppm, with a total range of 4.923 ppm. Station-wise, Red Deer-Riverside exhibits a slightly higher mean methane concentration (2.082 ppm) than Calgary Central (2.068 ppm). Both stations have similar standard deviations (0.222 ppm and 0.218 ppm, respectively), indicating comparable variability in methane levels.

2) *Percentile Distribution:* The entire dataset's interquartile range (IQR) is 0.235 ppm, with the 25th percentile (Q1) at 1.918 ppm and the 75th percentile (Q3) at 2.153 ppm, suggesting that the central 50% of the values are tightly clustered around the median. Station-specific percentile analysis reveals that Red Deer-Riverside has a wider IQR (0.300 ppm) as compared to Calgary Central (0.189 ppm), indicating greater variability in Red Deer-Riverside's measurements. Moreover, Calgary Central's 99th percentile reaches 2.900 ppm, higher than Red Deer-Riverside's 2.800 ppm, implying that Calgary Central experiences occasional extreme methane concentration levels.

TABLE I. DESCRIPTIVE STATISTICS OF METHANE CONCENTRATION (PPM)

Statistic	Calgary	Red Deer	Overall
Minimum	1.400	1.600	1.400
Q1	1.931	1.900	1.918
Mean	2.068	2.082	2.075
Median	2.000	2.000	2.000
Q3	2.120	2.200	2.153
Standard deviation	0.218	0.222	0.220
Maximum	6.323	4.600	6.323
Skewness	3.085	1.781	2.413
Kurtosis	21.405	6.808	13.742

3) *Distribution Characteristics:* The dataset exhibits a skewness of 2.413, implying a positively skewed distribution, and a kurtosis of 13.742, indicating a leptokurtic distribution with heavier tails. These characteristics suggest the presence of extreme values that could influence the performance of machine learning models, particularly in regression-based approaches. At the station level, Red Deer-Riverside's skewness is 1.781, with a kurtosis of 6.808. In contrast, Calgary Central has a higher skewness of 3.085 and kurtosis of 21.405, indicating that more extreme values exist in the Calgary Central dataset. The dataset's coefficient of variation (CV) is 10.620%, demonstrating relatively low variation in methane levels across the study period.

4) *Confidence Intervals:* The 95% confidence interval (CI) for the overall mean of methane concentration is [2.074, 2.076] ppm, suggesting a precise estimate of the expected methane levels. The station-specific confidence intervals are similarly narrow, with Red Deer-Riverside ranging between [2.080, 2.084] ppm and Calgary Central within [2.066, 2.070] ppm, further reinforcing the stability and reliability of the dataset. The station-specific confidence intervals are similarly narrow, with Red Deer-Riverside ranging inside [2.080, 2.084] ppm and Calgary Central within [2.066, 2.070] ppm, further reinforcing the stability and reliability of the dataset.

Figures 1 and 2 illustrate the distribution of methane concentrations across various monitoring stations and over multiple years.

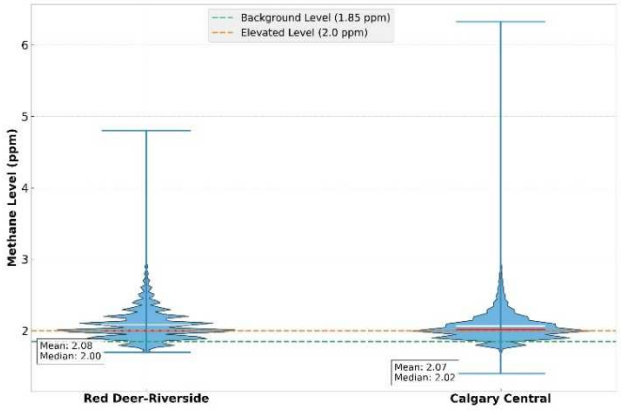

Fig. 1. Distribution of methane level by station.

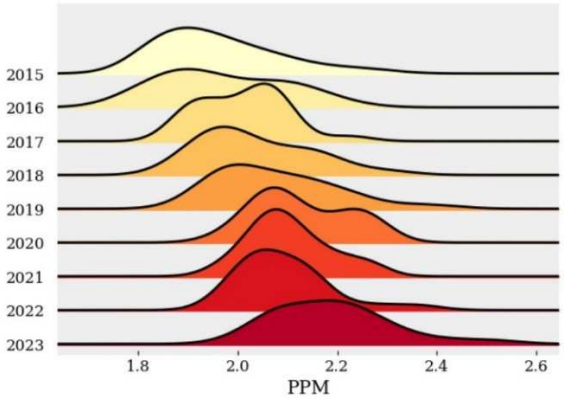

Fig. 2. Joyplot showing overall methane distribution over different years.

III. Machine Learning models

Machine learning is a field within computer science that allows systems to acquire knowledge from data and generate predictions without necessitating explicit programming. A key subfield of machine learning is deep learning, which uses multi-layered neural networks to model complex patterns. In time series, or temporal machine learning, models are designed to capture patterns and dependencies over time, making them useful for performing forecasting and trend analysis.

In this study, we utilized four different machine learning models for methane concentration prediction. Each model offers unique advantages in handling time series data—a data type that consists of observations recorded at regular, successive intervals and is widely used in disciplines such as economics, meteorology, environmental science etc. to allow researchers to analyze long-term trends, identify recurring patterns, and develop predictive models based on historical observations.

We split our dataset chronologically to maintain forecasting integrity. The training set contains data through December 31, 2022, the validation set spans the month of January in 2023, and the test set begins February 1, 2023. This temporal separation prevents future information from influencing past predictions. Additionally, transformations to a supervised learning format were applied independently per station to preserve the temporal and spatial integrity of the data. A brief discussion regarding the used machine learning models is as follows.

A. Random Forest

Random Forest (RF) [11] is a machine learning algorithm that works by creating a "forest" of decision trees. Each tree is trained on a random subset of the data, and the final prediction is made by averaging the predictions from all trees (for regression) or by taking the majority vote (for classification). It's like asking a group of experts for their opinions and using the consensus. It's known for its accuracy, robustness to overfitting, and ability to handle large datasets with many features.

B. XGBoost

XGBoost (Extreme Gradient Boosting) [12] is a gradient boosting algorithm that has gained popularity among machine learning practitioners. It works by combining multiple weak learners (usually decision trees) into a strong model. XGBoost is known for its high performance, scalability, and speed. It also includes regularization techniques to prevent overfitting, making it a powerful tool for both regression and classification tasks.

C. Multilayer Perceptron

A Multilayer Perceptron (MLP) [13] is a type of neural network that consists of multiple layers of nodes, each connected to the next. MLPs are called "feedforward networks" because information moves forward from the input to the output, without any loops. It's a versatile algorithm that can be used for tasks like classification and regression. While MLPs can be powerful, they require a good amount of data to train and can be more computationally intensive compared to tree-based models.

D. Temporal Convolutional Network

Temporal Convolutional Networks (TCNs) [14] are a type of deep learning model that is specifically designed for sequential data. Unlike traditional convolutional networks that work with spatial data, TCNs are tailored for time series data. They use a series of convolutions (filters) applied over time, which helps them learn patterns in data that evolve over time. TCNs are particularly useful for tasks like forecasting or predicting future events based on historical data.

IV. Analyses and Discussion

Before applying machine learning methods, we first analyzed the time series data by checking for autocorrelation, partial autocorrelation, and stationarity [15, 16] of the data. Autocorrelation helped us understand how current values are related to their past values, which can reveal repeating patterns or trends over time. Partial autocorrelation is built on this by showing the direct effect of past values at specific lags, removing the influence of the values in between. This made it easier to identify how many past time steps should be considered in the model. Based on these results, we found that a lag of around 3 was optimal, a result that was also confirmed using Optuna [17] with random forest model. Figures 3 and 4 show autocorrelation and partial autocorrelation plots.

We also performed the Augmented Dickey-Fuller (ADF) test to check if the data was stationary—meaning its statistical properties stay consistent over time. The test gave a p-value of 0.00, which clearly indicates the data is stationary and ready for modeling without further transformation.

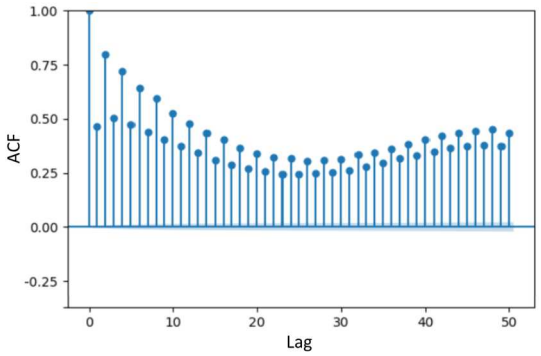

Fig. 3. Autocorrelation plot.

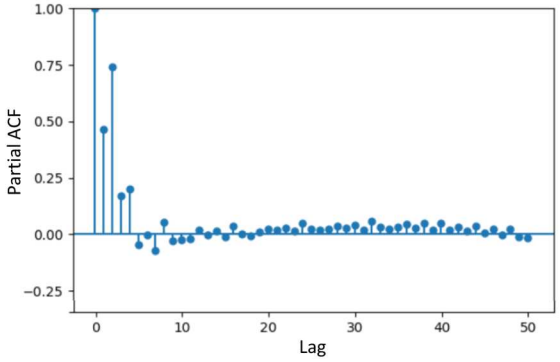

Fig. 4. Partial autocorrelation plot.

A. Model Equation using Machine Learning

The employed machine learning models predict the value at time *t* by considering the three previous time points of the target variable and one categorical feature. Since the categorical feature has two possible values, it is encoded as a single binary indicator. The prediction is represented as:

$$\widehat{y_t} = f(y_{t-1}, y_{t-2}, y_{t-3}, D_t) \quad (1)$$

here, $\widehat{y_t}$ is the predicted output, y_{t-1}, y_{t-2} and y_{t-3} are the values from the previous three time steps, and D_t is a dummy variable set to 0 or 1 (0 for Calgary Central, and 1 for Red Deer-Riverside). The function f symbolizes the machine learning model that captures how these inputs relate to the outcome.

The time series data was converted into a supervised learning dataset by using sliding window approach. The values from the previous three time steps were used as input features to predict the current value. By sliding this window across the data, we created input-output pairs that allow the model to learn patterns over time for better forecasting.

B. Performance Metrics

When developing models to predict numerical values, it is important to use performance metrics to evaluate the accuracy of predictions. In this study, five key performance metrics were used to assess how effectively the employed machine learning models performed in the regression analysis: the coefficient of determination (R^2), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Relative Root Mean Square Error (RRMSE), and Mean Absolute Percentage Error (MAPE) [18, 19].

R^2 (coefficient of determination) shows how much of the variation in the actual data is explained by the model. It ranges from 0 to 1, and a value closer to 1 means the model fits the data well. RMSE (Root Mean Square Error) measures the average size of the prediction errors, giving more weight to large mistakes. MAE (Mean Absolute Error) also measures average error, but treats all mistakes equally, making it easier to understand and less affected by extreme values. RRMSE (Relative RMSE) is the RMSE divided by the average of the actual values, which helps compare models across different datasets or scales. MAPE (Mean Absolute Percentage Error) shows the average error as a percentage, making it easy to see the divergence between the actual and predicted values in relative terms. Together, these metrics provide a comprehensive view of the model's accuracy and reliability. The mathematical equations for all five performance metrics are given below.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \quad (2)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2} \quad (3)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right| \quad (4)$$

$$RRMSE = \frac{\frac{1}{n}\sqrt{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}}{\bar{Y}} \times 100 \quad (5)$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \hat{Y}_i}{Y_i}\right| \times 100 \quad (6)$$

where Y_i = actual value, $\hat{Y}_i$ = predicted value, $\bar{Y}$ = mean of actual values, n = number of data points.

TABLE II. PERFORMANCE METRICS

Model	R^2	RMSE	MAE	RRMSE	MAPE
Random Forest	0.809	0.089	0.045	3.510	1.930
XGBoost	0.806	0.090	0.045	3.550	1.950
MLP	0.810	0.089	0.044	3.530	1.910
TCN	0.813	0.088	0.044	3.490	1.890

Table II provides test data performance results for all the employed models. The results show that all four models gave good predictions, with only small differences between them. The Temporal Convolutional Network (TCN) demonstrates the most reliable performance. It scores the highest in terms of R^2 (0.813), suggesting it captures the most variance in the target variable. Additionally, it achieves the lowest error rates across all other metrics, including RMSE (0.088), MAE (0.044), RRMSE (3.490), and MAPE (1.890), indicating strong predictive accuracy and consistency. While MLP performs almost on par with TCN—particularly in terms of MAE and MAPE—Random Forest and XGBoost show slightly less favorable results, with XGBoost yielding the lowest R^2 and the highest error values. Overall, TCN stands out as the top-performing model, especially in applications involving sequential or time-dependent data, with MLP being a solid secondary choice. Although the differences are minimal, the neural network-based models (TCN and MLP) had a slight edge over the others in this case.

C. Shapley Additive Explanations

Shapley Additive Explanations (SHAP) [20] is used to interpret machine learning models by assigning each feature a contribution value for a particular prediction. Based on game theory, SHAP provides a consistent and fair way to understand how individual features influence a model's output, making complex models like random forests and neural networks more transparent and explainable.

The SHAP analysis revealed that previous time step, value(t-1), stood out with the highest importance score of 0.1529, showing that recent values in the time series are key drivers in predicting the target. Value(t-2) also contributed, though its impact was much smaller (0.0149). Among the station-related features, Red Deer-Riverside and Calgary Central showed modest importance, while value(t-3) had the lowest contribution. Overall, the results emphasize that the model relies heavily on recent data points, confirming the temporal nature of the information. Figure 5 displays a graphical representation of the SHAP results for the best performing model i.e., TCN.

Lastly, Fig. 6 shows how well the TCN model predicted methane levels compared to the actual values, highlighting its excellent performance. TCN outperformed the other models,

likely because it captured long-term patterns in the time series more effectively and trained more efficiently

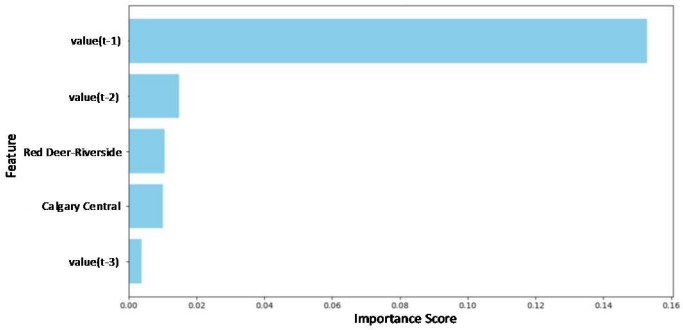

Fig. 5. SHAP summary plot for TCN model.

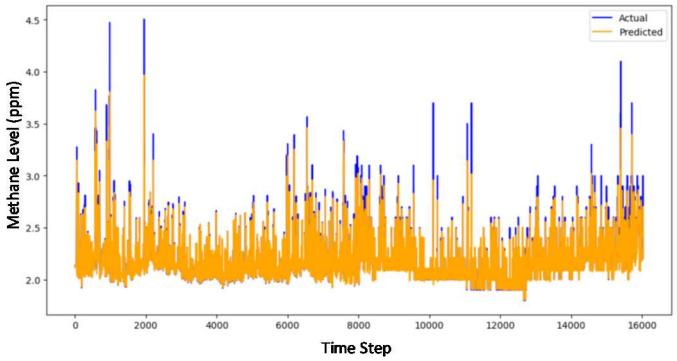

Fig. 6. Actual and predicted methane level using TCN model.

V. CONCLUSION

This work used Alberta's methane warehouse time series data to analyze the performance of different machine learning algorithms and demonstrate the strength they provide to carry out robust predictive analysis. In summary, all four machine learning models namely Random Forest, XGBoost, MLP, and TCN exhibited strong performance in forecasting the output. The TCN emerged as the best performer, attaining the highest R^2 value (0.813), which indicates the strongest ability to explain the variance in the target data. It also recorded the lowest error values across all remaining metrics, including RMSE (0.088), MAE (0.044), RRMSE (3.490), and MAPE (1.890), highlighting its superior accuracy and consistency in predictions. The Multilayer Perceptron (MLP) subsequently showed marginally worse performance measures. Both Random Forest and XGBoost yielded commendable results, but somewhat lower to the neural network-based models.

SHAP was employed to elucidate the contribution of each characteristic to the predictions, hence improving the interpretability of the results. This showed the predominant elements in the time series data and offered a simpler comprehension of the models' predictive mechanisms. In summary, although the distinctions among the models were negligible, the neural network models (TCN and MLP) provided a marginal advantage, especially with their capacity to capture the temporal dynamics within the data.

This work clearly shows that temporal machine learning methods can be a practical way to handle time series prediction problems and can complement traditional approaches. What makes this study stand out is the use of advanced models like the TCN for predicting methane discharge, which proved especially effective at capturing complex patterns in the data.

Although, this study exhibits notable gains, our future actions involve expanding these accomplishments by incorporating recurrent neural network (RNN) or employing a hybrid model to further improve the prediction results. Additionally, more monitoring stations will be added to broaden the study to a more complex dataset. Residual plots for real versus predicted time series will be added to illustrate trends. Statistical significance tests will be undertaken to assess if variations between model performances are relevant or not.

VI. ACKNOWLEDGMENTS

The authors would like to express their gratitude to the Department of Information Systems and Technologies (CTIS) at Bilkent University for their continuous academic and administrative support throughout this research. Researchers also extend their appreciation to the University for providing partial financial support, which played an important role in making this work possible.

REFERENCES

[1] Climate and Clean Air Coalition, "Methane: The Global Methane Pledge," CCAC, 2021.

[2] Intergovernmental Panel on Climate Change, *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the IPCC*, Cambridge University Press, 2021.

[3] National Oceanic and Atmospheric Administration, "Increase in atmospheric methane set another record during 2021," NOAA, 2022.

[4] A. R. Van Amstel, "Methane: A review," *Journal of Integrative Environmental Sciences*, vol. 9, no. 1, pp. 5-30, 2012.

[5] D. Shindell et al., "Simultaneously mitigating near-term climate change and improving human health and food security," *Science*, vol. 335, no. 6065, pp. 183-189, 2012.

[6] D. Güleryüz, "Forecasting of methane (CH4) gas emission through grey model, ARIMA and exponential smoothing: the case of Turkey," *Environmental Science and Pollution Research*, vol. 30, pp. 12345-12367, 2023.

[7] Y. Luo et al., "A novel hybrid model for methane concentration prediction in Alberta, Canada," *Environmental Science and Pollution Research*, vol. 31, pp. 54321-54345, 2024.

[8] M. Aslan et al., "Prediction of methane concentration using machine learning methods: the case of Turkey,"*Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, vol. 43, no. 20, pp. 2505-2521, 2021.

[9] Government of Alberta, "Alberta Air Data Warehouse," [Online]. Available: https://environment.alberta.ca/apps/EdwReportViewer/StationInformationAndStatus.aspx [Accessed: 08-05-2025].

[10] Yoon, J., Zame, W., & van der Schaar, M. (2018). Estimating Missing Data in Temporal Data Streams Using Multi-Directional Recurrent Neural Networks. 2018 IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 1275–1280. doi: 10.1109/ICMLA.2018.00206

[11] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.

[12] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 2016, pp. 785-794.

[13] J. R. Parker, "Multilayer perceptrons," IEEE Transactions on Neural Networks, vol. 1, no. 1, pp. 56-67, 1990.

[14] L. Bai, J. Z. Kolter, and V. T. Lee, "Temporal convolutional networks for action segmentation and detection," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 2018, pp. 1-11.

[15] R. H. Shumway and D. S. Stoffer, Time Series Analysis and Its Applications: With R Examples, 4th ed., Springer, 2017.

[16] J. D. Hamilton, Time Series Analysis, Princeton University Press, 1994.

[17] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," *Proc. 25th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, Anchorage, AK, USA, 2019, pp. 2623-2631, doi: 10.1145/3292500.3330701.

[18] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York, NY, USA: Springer, 2009.

[19] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning: with Applications in R*. New York, NY, USA: Springer, 2013.

[20] S.M. Lundberg, & S.I. Lee, (2017). A unified approach to interpreting model predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS 2017)*, 4768-4777. https://doi.org/10.1145/3295222.3295231