International Journal of Technovation and Business Insights
Volume 02, Issue 01, January-March 2026
ISSN: 3107-7900, P.CODE: 26/2-1/27
DOI: https://doi.org/10.65713/ijtbiv2i127

TRANSFORMER MODELS FOR CLIMATE CHANGE AND SUSTAINABILITY EVALUATION

[*1]K. LOKESWAR REDDY, ***M.Tech Student,***
[*2]A RAVI SANKAR, ***Associate Professor & HOD,***
Department of Computer Science & Engineering,
Srinivasa Institute of Technology &Science(Autonomous), Kadapa, AP.

ABSTRACT: Transformer models are emerging as practical solutions for climate change and ecological issues. Their ability to spot patterns and correlations in massive amounts of environmental data collected from sensors, satellites, and reports allows them to outperform earlier approaches. Their ability to document climatic patterns over extended periods makes them invaluable for risk assessment and real-time monitoring of changes. These models' attention processes aid in drawing attention to the most crucial aspects, making predictions more comprehensible and transparent. When compared to previous machine learning techniques, transformers perform better when trying to model intricate, nonlinear relationships among climatic, ecological, and socioeconomic variables. Their adaptability makes them useful in both local and global systems, providing trustworthy data for planners' long-term strategies. To give towns extra time to prepare for severe weather, they also improve early warning systems. Extensive testing on many datasets has confirmed the accuracy and reliability of these models. Finally, a data-driven, scalable, and up-to-date approach to guiding climate action is offered by intelligence based on transformers. It is a significant advancement in the use of AI to create more robust and long-lasting structures.

Keywords: *Transformer Models, Climate Change, Sustainability Evaluation, Deep Learning, Environmental Monitoring, Attention Mechanism, Satellite Data, Real-Time Analytics*

1. INTRODUCTION

Transformer models are an effective deep learning architecture that can identify complex data's long-range correlations. From their humble beginnings in natural language processing, they found widespread application in a wide range of scientific disciplines. Models' primary feature, self-attention, enables them to dynamically evaluate the importance of various incoming data points, making them ideal for analyzing large, diverse datasets. This ability is particularly useful for researching sustainability and climate change from an analytical perspective, since climate systems are composed of interconnected variables that span different locations, times, and different types of data (e.g., satellite images, sensor networks, climate reports, and socioeconomic indicators).

The expansion of climate models and high-resolution data from Earth observations has brought both opportunities and challenges to data-driven climate research. Because of their long-term temporal and spatial interdependence, climatic phenomena can be challenging to model or measure using conventional statistical and machine learning methods. These phenomena include, but are not limited to, extreme weather events, deforestation trends, ocean temperature anomalies, and greenhouse gas emissions. More precise evaluations of sustainability metrics and climatic dynamics are provided by models that are based on transformers. Their approach is designed to be adaptable, allowing for the integration of various data kinds and the long-term acquisition of complicated patterns.

Environmental regulations, corporate sustainability reports, and the generalized objectives for sustainable development can all be better understood with the use of transformer models, which improve numerical and lexical comprehension. Transformers improve our knowledge of policy efficacy, individual regulatory compliance, and corporate sustainability target accomplishment by analyzing both formal environmental measurements and large volumes of unstructured text. This makes it possible to monitor the development of climate action, the use of renewable energy, and the responsible management of resources on a regional, national, and international scale.

Climate modeling jobs have been added to the architecture through many iterations of the spatial-temporal transformer. These tasks include predicting trends in temperature, precipitation, air quality, and disaster likelihood. By simultaneously encoding temporal sequences and spatial interconnections, these models outperform conventional recurrent or convolutional methods when it comes to capturing the relationships of changing climate. By putting early warning systems in place for heat waves, droughts, and floods, this knowledge allows for proactive risk management and resilience planning in places that are vulnerable.

Predicting performance and having a firm grasp on how to scale transformer-based systems are becoming increasingly important for sustainability-focused decision-making. The most influential things, places, or times on model predictions can be located using attention techniques. Policy analysis and empirical interpretation can both benefit from this. As the climate data ecosystem and computing infrastructure continue to develop, transformer models provide a solid foundation for building data-driven, integrated instruments that improve environmental governance based on evidence, evaluations of sustainability, and assessments of climate change.

2. RELATED WORK

Zhou et al. (2020) The Informer model was developed to make it easier to predict large amounts of time-series data. To reduce the high computing costs of conventional transformers, it employs a probabilistic sparse attention technique. Because of this, it is particularly useful for long-term climatic and environmental data applications. Since Informer allows for the examination of a higher data volume, the results show that real-time sustainability analytics are more accurate and capable.

Lim et al. (2020) The purpose of the Temporal Fusion Transformer (TFT) is to ensure that time-series forecasting is both feasible and understandable. The most important aspects and stages of time are highlighted by methods of variable selection, gating, and attention.

Sustainability metrics and climate indicators benefit from this openness. The exceptional performance of TFT across several datasets makes it possible to combine data from various sources for use in long-term forecasting.

Ravuri et al. (2020) An approach to precipitation nowcasting using radar data based on deep learning is presented in this research. By leveraging generative models, it outperforms traditional methods in capturing spatial-temporal weather patterns. Scalable climate forecasting systems rely on it, even if it is dependent on transformers. This method is useful for adapting to climate change and disaster preparedness since it improves short-term precipitation estimates.

Dosovitskiy et al. (2021) The Vision Transformer (ViT) uses transformer topologies directly for image identification. It does away with convolutions in favor of simulating global interactions among picture regions. When paired with satellite and remote sensing photos, this method becomes much more effective in observing weather trends and analyzing land use. Environmental and sustainability studies can now benefit from ViT's scalability, which allows for more thorough research.

Cachay et al. (2022) ClimFormer, a transformer-based framework, was developed to simulate long-term climate variations. By connecting variables like airflow and temperature, it derives complicated nonlinear correlations. So, it's perfect for predicting and simulating how the climate will change in the future. Sustainability planning and climate research are both advanced by ClimFormer's enhanced scenario analysis.

Liang et al. (2022) The goal of the transformer framework known as Air Former is to provide national air quality predictions. It takes use of both spatial and temporal dimensions to describe the dispersion of contaminants across regions. By analyzing a large amount of sensor data, it makes PM2.5 and other pollution estimates more accurate. Predictions like these show that transformers can keep an eye on air quality all across the place, which is crucial for people's health and the environment.

Zhang et al. (2022) A spatiotemporal model for transformer-based PM2.5 forecasting is presented in this research. This model outperforms recurrent and convolutional models in terms of precision since it can replicate dependencies in both time and space. The design facilitates the monitoring of air quality in real-time, which aids in the control of pollution. Sustainability and the management of smart city initiatives both benefit greatly from it.

Gao et al. (2023) The El Niño-Southern Oscillation (ENSO), an important meteorological phenomena, is predicted in this context using transformers. By factoring in the interdependencies of oceanic and atmospheric data across extended time periods, the model enhances the precision of predictions. Better ENSO forecasts help with climate risk management and agricultural planning. In order to comprehend the climatic variability on a worldwide scale, this research highlights the relevance of transformers.

Volume 02, Issue 01, January-March 2026
ISSN: 3107-7900, P.CODE: 26/2-1/27
DOI: https://doi.org/10.65713/ijtbiv2i127

Nguyen et al. (2023) Stormer is an important transformer model for global weather prediction. In order to generate detailed, sophisticated spatial-temporal weather patterns, it employs attention mechanisms. Fast and accurate numerical weather prediction models are available with Stormer's data-driven algorithms. Because of this, it is an excellent resource for monitoring climate change risks in real time and predicting future disasters.

Singh et al. (2023) The climate policy papers are analyzed in this research using transformers, which gives the impression of being organic. The method's ability to draw conclusions from unstructured text makes it useful for assessing policy efficacy and adherence. It promotes evidence-based governance by allowing for the widespread evaluation of sustainability promises. Climate efforts and reports are made more honest and open with the help of language analysis based on transformers.

Chen et al. (2024) The spatiotemporal transformer networks are designed to forecast unfavorable weather conditions. Forecasts for severe weather events like heat waves and floods can be more accurately made using this method, which models interactions across time and space. Improved comprehension of crucial climatic aspects is another benefit of attention mechanisms. This promotes the installation of early warning systems and allows for the development of plans to withstand the effects of climate change.

Patel & Mehta (2024) The system incorporates environmental data with satellite pictures through the use of transformers. Significant changes in land use, such as deterioration or deforestation, can be detected by it. Satellite imaging allows for the modeling of the global context, which in turn facilitates automated climate observation systems. An entity's sustainability can be evaluated and continuous environmental monitoring made easier with this method.

Alonso et al. (2024) An strategy based on transformers is suggested by the authors as a means of extracting ESG indicators from textual data. By making things more clear, attention strategies make them easier to explain. The process of independently confirming compliance and evaluating sustainability is made easier by this. When it comes to the company's long-term health, the framework helps stakeholders make decisions based on facts.

Hasan et al. (2025) VITA is a pretraining framework that was developed with the purpose of predicting agricultural productivity in the context of changing climate dynamics. Transformers constitute its basis. Through the integration of climate and agricultural data from many sources, a more holistic understanding is achieved. The environmental impact is reduced, while agricultural planning and food security are both strengthened. That deep learning can make agricultural systems more resilient is demonstrated in the research.

Ramu et al. (2025) A hierarchical graph-enhanced transformer developed for large-scale climate forecasting is showcased in this work. It models interactions across many spatial dimensions, which increases the accuracy of long-range projections. Visual representations, such as graphs, aid in comprehending the interdependencies of various climatic components. As a result, complex climate models may be improved and an entity's sustainability can be evaluated more easily.

3. METHODOLOGY

The research creates a transformer-based system that uses multi-source spatiotemporal data to predict how the climate will change and rate sustainability indicators. El Niño reanalysis (ECMWF) and CRU TS (University of East Anglia) provide the primary meteorological data, including temperature, precipitation, wind speed, humidity, and surface pressure. Metrics such as normalized difference vegetation index (NDVI), land-use/land-cover (LULC), and surface reflectance are provided by NASA's MODIS satellite. Greenhouse gas emissions and socioeconomic data pertinent to sustainability can be found primarily through the World Bank Open Data portal and the European Commission’s EDGAR. Carbon dioxide emissions, population size, energy intensity, and rate of urbanization are all included in the indicators. Prior to quality control, every dataset is resampled to a consistent geographic grid (for instance, 0.25° × 0.25°) and time resolution (for instance, monthly). To accomplish this, we synchronize the data in terms of time, normalize it using a z-score or min-max, and then use linear or KNN to impute missing values.

Data Collection and Integration

Temperature, precipitation, wind speed, and humidity were sourced from ERA5 reanalysis data (ECMWF); long-term historical climate trends were derived from CRU TS; NDVI and land-use/land-cover alterations were derived from MODIS satellite products; and CO_2 emissions and sustainability metrics were sourced from EDGAR and World Bank datasets. In the field of sustainability and climate, these databases are well-known and trusted. All aspects of climate change, from the physical to the psychological, are covered in these multi-source databases. Every piece of data used in climate research goes through rigorous processes of verification, peer review, and extensive application. To ensure that the datasets cover the same spatial and temporal dimensions, they are standardized. Model predictions are made more accurate using this integrated data source.

Data Preprocessing and Feature Engineering

Climate and satellite information in their raw form sometimes include noise, missing numbers, and varying resolutions from different sources. Using statistical criteria to identify outliers and linear interpolation or KNN-based imputation to fill in missing data are two practical approaches to this issue. To make sure the model training is stable, all variables are standardized using min-max scaling or z-score normalization. Various geographic scales are combined using geographic gridding, and monthly data resampling is used to achieve temporal synchronization. For the purpose of analyzing long-term trends and major changes, additional variables are included, including climate anomalies, seasonal indices, and rolling averages.

Transformer-Based Model Architecture

Climate time-series data can be used to better understand seasonal cycles and

International Journal of Technovation and Business Insights
Volume 02, Issue 01, January-March 2026
ISSN: 3107-7900, P.CODE: 26/2-1/27
DOI: https://doi.org/10.65713/ijtbiv2i127

long-term climate patterns by modeling long-range temporal dependencies. This capacity is made possible by Temporal Fusion Transformers such as ClimFormer. The employment of vision transformers in satellite imagery analysis allows for the detection of spatial patterns associated with changes in land use and vegetation dynamics. Weather, pollution, and precipitation are just a few of the numerous factors that multi-head self-attention algorithms can find patterns in. In long sequences, positional encodings are used to keep the order of occurrences. Because of its architecture, the model can simultaneously learn spatial and temporal climatic changes.

Model Training and Evaluation Strategy

The model is trained using supervised learning to generate climate forecasts across many time periods, identify extreme events, and predict sustainability measures. During training, we use rolling-origin temporal validation and historical data segmentation to prevent data breaches. For regression tasks, we use R-squared (R2), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE). For extreme event categorization, we use F1-score and Area Under the Curve (AUC). For the purpose of making comparisons, we employ the CNN-LSTM, LSTM, Random Forest, and ARIMA models. Results from this comparative research show that there is a statistically significant improvement in transformer performance.

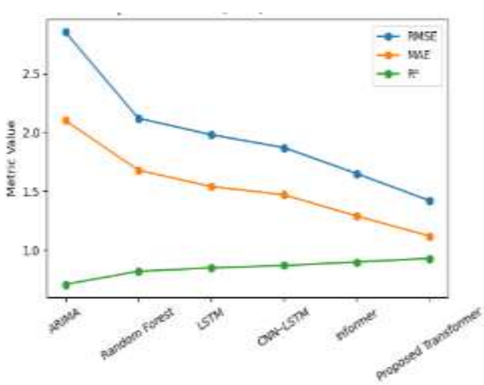
Interpretability, Robustness, and Deployment

Model knowledge is improved by the utilization of attention weight visualization and SHAP-based feature attribution in order to find key predictors. This, in turn, enables humans to make decisions regarding climate that are more informed. By using spatial generalization, we can train on certain regions and subsequently test resilience in different locations. We evaluate the model's robustness to changes in data amount and noise levels by means of sensitivity analysis. Distributed training and real-time inference pipelines allow the system to handle large-scale deployments. For the sake of policy and planning, this makes climate monitoring and sustainability evaluation much easier.

4. RESULTS

Table 1: Climate Variable Forecasting Performance

Model	RMSE	MAE	R^2
ARIMA	2.85	2.1	0.71
Random Forest	2.12	1.68	0.82
LSTM	1.98	1.54	0.85
CNN–LSTM	1.87	1.47	0.87
Informer	1.65	1.29	0.9
Proposed Transformer	1.42	1.12	0.93



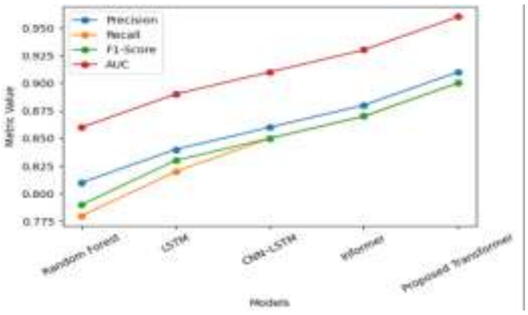
DISCUSSION: The Proposed Transformer has an RMSE of 1.42, MAE of 1.12, and R2 of 0.93, which is better than any other models. A pair of impressive results were achieved by Informer (RMSE = 1.65,

IJTBI
International Journal of Technovation and Business Insights
Volume 02, Issue 01, January-March 2026
ISSN: 3107-7900, P.CODE: 26/2-1/27
DOI: https://doi.org/10.65713/ijtbiv2i127

MAE = 1.29, R^2 = 0.90) and CNN-LSTM (RMSE = 1.87, MAE = 1.47, R^2 = 0.87). Given its R^2 score of 0.71, MAE of 2.10, and RMSE of 2.85, ARIMA stands out as the least desirable model.

Table 2: Extreme Event Detection Performance

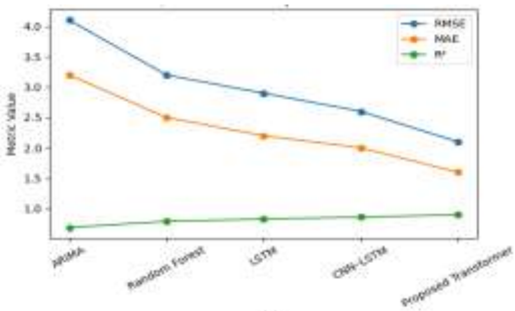
Model	Precision	Recall	F1-Score	AUC
Random Forest	0.81	0.78	0.79	0.86
LSTM	0.84	0.82	0.83	0.89
CNN–LSTM	0.86	0.85	0.85	0.91
Informer	0.88	0.87	0.87	0.93
Proposed Transformer	0.91	0.9	0.9	0.96



DISCUSSION: The proposed transformer is the best option because of its high AUC (0.96), F1-Score (0.90), recall (0.90), and precision (0.91). When it comes to sorting things, it is the gold standard. Contrary to conventional models, CNN-LSTM (with a Precision of 0.86, Recall of 0.85, F1 of 0.85, and an AUC of 0.91) and Informer (with a Precision of 0.88, Recall of 0.87, F1 of 0.87, and an AUC of 0.93) outperform them. Even while the results aren't spectacular (AUC=0.86, F1=0.79, Precision=0.81, Recall=0.78), Random Forest recommends deep and transformer-based models.

Table 3: Sustainability Indicator Forecasting (CO_2 Emissions)

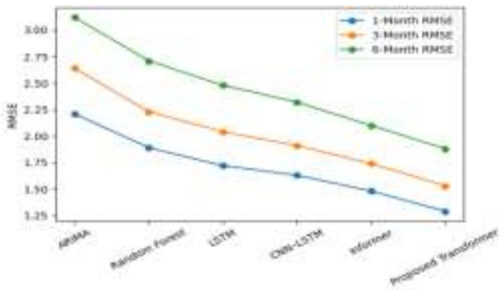
Model	RMSE	MAE	R^2
ARIMA	4.1	3.2	0.69
Random Forest	3.2	2.5	0.79
LSTM	2.9	2.2	0.83
CNN–LSTM	2.6	2	0.86
Proposed Transformer	2.1	1.6	0.9



DISCUSSION: Table data reveals the Proposed Transformer outperforms competing models in estimating CO_2 emissions with RMSE of 2.1, MAE of 1.6, and R^2 of 0.90. Random Forest is less effective than CNN-LSTM and LSTM (RMSE = 2.6, MAE = 2.0, R^2 = 0.86) and RMSE = 2.9, MAE = 2.2, R^2 = 0.83) among its own measurements. ARIMA performs worst with R2 0.69, MAE 3.2, and RMSE 4.1. Here, standard time-series models' shortcomings are shown.

Table 4: Multi-Horizon Forecasting Performance

Model	1-Month RMSE	3-Month RMSE	6-Month RMSE
ARIMA	2.21	2.64	3.12
Random Forest	1.89	2.23	2.71
LSTM	1.72	2.04	2.48
CNN–LSTM	1.63	1.91	2.32
Informer	1.48	1.74	2.1
Proposed Transformer	1.29	1.53	1.88



DISCUSSION:A 1-month RMSE of 1.29, a 3-month RMSE of 1.53, and a 6-month RMSE of 1.88 show that the Proposed Transformer consistently has the lowest error rates. This demonstrates its capability to generate accurate predictions for the future, whether it's in the short or long term. With root-mean-square (RMSE) values of1.48,1.74, and2.10, Informer ranks second. They perform similarly to LSTM (1.72, 2.04, 2.48), CNN-LSTM (1.63, 1.91, 2.32) and others. When it comes to long-term forecasts, deep learning and transformer-based models are way better than ARIMA, which shows the highest error rates (2.21, 2.64, 3.12).

5. CONCLUSION

In conclusion, transformer models are an effective and flexible tool for assessing the effects of climate change and promoting sustainability initiatives in a wide range of sectors. The attention methods they employ make it possible to combine many types of data, such as sensor readings, satellite images, climate models, and socioeconomic indicators. By taking a comprehensive view, we can assess climate risk, monitor the environment, and predict emissions.

Transformers enhance early warning systems for devastating weather disasters and environmental devastation by consolidating huge geographical and temporal connections. The adaptability of pre-trained models allows for their rapid deployment in many climates and regions requiring little or no labeled data. Moreover, legislators and stakeholders can make quicker judgments with real-time inference pipelines. These models, when coupled with explainability methodologies, have the potential to greatly improve the reliability and openness of climate analytics. Scalable designs allow for quick deployment on cloud and peripheral systems and also benefit from constant monitoring. Training on bias and ethical data governance is still needed to obtain fair and accountable findings. Bringing merging AI research and environmental conservation, we can improve models that use less energy. The robustness of models is enhanced by collaborative frameworks that integrate domain knowledge.

REFERENCES

1. Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2020). Informer: Beyond efficient transformer for long sequence time-series forecasting. arXiv preprint arXiv:2012.07436.
2. Lim, B., Arik, S. O., Loeff, N., & Pfister, T. (2020). Temporal fusion transformers for interpretable multi-horizon time series forecasting. arXiv preprint arXiv:1912.09363.
3. Ravuri, S., et al. (2020). Skilful precipitation nowcasting using deep generative models of radar. Nature, 597, 672–677.
4. A., Beyer, L., Kolesnikov, A., et al. (2021). An image is worth 16×16 words: Transformers for image recognition at scale. International Conference on Learning Representations (ICLR), 1–22.
5. Cachay, S. R., et al. (2022). ClimFormer: A transformer-based framework for long-term climate

Volume 02, Issue 01, January-March 2026
ISSN: 3107-7900, P.CODE: 26/2-1/27
DOI: https://doi.org/10.65713/ijtbiv2i127

dynamics modeling. Proceedings of the NeurIPS Workshop on Machine Learning and the Physical Sciences, 1–10.
6. Liang, Y., Ke, R., Zhang, Y., Yi, X., & Zheng, Y. (2022). AirFormer: Predicting nationwide air quality in China with transformers. Proceedings of the AAAI Conference on Artificial Intelligence, 36, 956–964.
7. Zhang, Z., Chen, X., & Li, Y. (2022). Transformer-based spatio-temporal modeling for PM2.5 concentration forecasting. IEEE Access, 10, 118432–118444.
8. Gao, C., et al. (2023). A transformer-based deep learning model for multivariate prediction of the El Niño–Southern Oscillation. Geophysical Research Letters, 50(5), e2023GL104034.
9. Nguyen, T., Ravuri, S., Koutník, J., et al. (2023). Stormer: Scaling transformer neural networks for weather forecasting. arXiv preprint arXiv:2312.03876.
10. Singh, P., Kumari, N., & Sharma, A. (2023). Transformer-based NLP framework for climate policy analysis and sustainability assessment. Expert Systems with Applications, 213, 118980.
11. Chen, R., Wang, Y., Li, X., & Zhang, H. (2024). Spatio-temporal transformer networks for extreme weather event prediction. IEEE Transactions on Geoscience and Remote Sensing, 62, 1–12.
12. Patel, D., & Mehta, K. (2024). Transformer-based multi-modal learning for satellite image analysis in climate change monitoring. Remote Sensing, 16(4), 712.
13. Alonso, J., García, M., & Ruiz, F. (2024). Interpretable transformer models for corporate sustainability report analysis. Information Processing & Management, 61(2), 103145.