

Integrating Indigenous Knowledge with Meteorological Data for Explainable Climate Prediction: An LSTM-SHAP Framework for Archipelagic Regions



Iis Hamsir Ayub Wahab^{1*}, Mohamad Jamil², Muhammad Said¹

¹ Department of Electrical Engineering, Universitas Khairun, Ternate 97719, Indonesia

² Department of Informatics Engineering, Universitas Khairun, Ternate 97719, Indonesia

Corresponding Author Email: hamsir@unkhair.ac.id

Copyright: ©2026 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/isi.310407>

ABSTRACT

Received: 9 November 2025

Revised: 20 January 2026

Accepted: 11 April 2026

Available online: 30 April 2026

Keywords:

Indigenous Knowledge Systems, Explainable Artificial Intelligence, climate prediction, Long Short-Term Memory, SHapley Additive exPlanations, knowledge integration, Archipelagic climate

Bridging scientific climate data with Indigenous Knowledge Systems (IKS) remains methodologically challenging, particularly in data-sparse archipelagic regions where conventional meteorological observations are limited. This study proposes an Explainable Artificial Intelligence (XAI) framework that integrates IKS-based ecological indicators with meteorological variables for interpretable climate prediction. Indigenous indicators—including leaf curling, cloud formation over mountain peaks, seawater turbidity, and bioluminescent plankton—were systematically collected through ethnographic fieldwork in North Maluku, Indonesia, and encoded as structured features. A bidirectional Long Short-Term Memory (BiLSTM) model was trained on multivariate time-series data (12 locations, 450 days) combining five scientific variables (rainfall, temperature, humidity, wind direction) and five IKS-derived indicators to predict daily rainfall events. The model achieved a recall of 99.79% for rainfall detection, demonstrating high sensitivity suitable for risk-averse early warning applications, with an overall accuracy of 51.32% and precision of 51.37%, reflecting the impact of severe class imbalance and simplified binary encoding of indigenous indicators. SHapley Additive exPlanations (SHAP) analysis revealed that while meteorological variables (rainfall, humidity, temperature) dominate predictions, IKS indicators—particularly cloud cover over mountains and leaf curling—show consistent positive contributions aligned with their traditional ecological interpretations. Rather than pursuing high predictive accuracy, this study contributes a transparent, empirically grounded methodology for operationalizing and interpreting Indigenous Knowledge within data-driven climate analysis. The framework demonstrates that XAI can serve as an interpretive bridge between scientific and indigenous epistemologies, enabling semantically aligned knowledge integration. These findings support the development of context-aware, interpretable climate services that respect and incorporate local knowledge systems.

1. INTRODUCTION

Climate change is a growing, complex global threat that impacts ecological systems and human life worldwide. In addressing this crisis, climate prediction models play a vital role in providing the scientific information needed for adaptation and mitigation planning. However, a significant challenge often faced is the limited interpretability of these models. Many scientific approaches based on Artificial Intelligence (AI), although accurate, are black-box in nature and difficult for stakeholders, including communities directly affected by climate change, to understand.

On the other hand, for centuries, various indigenous communities have developed rich and adaptive knowledge systems for reading the signs of nature and managing the environment sustainably [1, 2]. These systems are known by various terms such as Indigenous Knowledge Systems (IKS) [3] or Traditional Ecological Knowledge (TEK) [4], each emphasizing the deep relationship between humans and their environment. Broadly, IKS represent generations of localized

environmental understanding and community-based adaptation practices. Rather than focusing solely on ecological prediction, IKS provide holistic frameworks that integrate social, cultural, and institutional strategies to enhance community resilience and inform climate change education [5]. Research such as that conducted by Motsumi and Nemaconde [6] confirm that early warning systems grounded in IKS are highly practical and context-specific, enabling rural communities to anticipate hazardous events through environmental indicators such as vegetation changes, lunar cycles, and animal behavior. They emphasize that integrating these localized early warning indicators with meteorological forecasts can strengthen disaster preparedness and enhance community resilience [6]. Furthermore, research conducted by Jiri et al. [7] shows that although traditional knowledge still holds considerable predictive value, many natural indicators that were once considered stable such as seasonal wind cycles, animal behavior, and rainfall patterns have become increasingly unreliable due to the impacts of ongoing climate variability and change [7]. Similarly, Mafongoya et al. [8]

revealed that the validity and sustainability of IKS are under significant pressure due to multiple factors, including the influence of Western-oriented formal education systems that tend to marginalize local knowledge, the growing dependence on scientific seasonal forecasting (SSF), and the impacts of climate change that disrupt the stability of traditional ecological indicators [8].

These conditions underscore the urgency of formulating an integrative approach that can accommodate both modern scientific knowledge and local ecological wisdom. Instead of dichotomizing the two, a transdisciplinary framework is required to build epistemological synergy. In this regard, Explainable Artificial Intelligence (XAI) provides a promising paradigm that not only enhances predictive accuracy but also ensures transparency, interpretability, and accountability in model decisions [9-11].

Through the XAI approach, this study aims to develop a climate prediction model that is not only technically sophisticated but also inclusive and communicative. By bridging scientific and IKS, it is hoped that a new paradigm in climate modeling will emerge that is not only based on data and algorithms but also on human values, contexts, and experiences. This approach is a strategic step in creating climate change solutions that are equitable, contextual, and sustainable.

2. RELATED WORKS

Various studies show that the application of XAI in climate and environmental science is increasingly important to address the “black box” nature of complex AI models and to increase public and policymakers' confidence in prediction results. A comprehensive review by Huang et al. [12] highlights that XAI in Earth System Science not only improves the interpretability and transparency of AI models but also serves as a scientific tool to uncover dominant patterns and physical relationships underlying climate processes, effectively bridging the gap between process-based and data-driven approaches [12]. Meanwhile, in their systematic literature review, Schiller et al. [13] analyzed 575 publications on the use of XAI in environmental and Earth system sciences. They found that although methods such as SHAP and LIME are widely applied, few studies systematically evaluate the faithfulness and robustness of explanations. This indicates the need for standardized evaluation frameworks to ensure the reliability and trustworthiness of XAI applications in complex environmental systems [13].

In the context of explainable climate prediction, some studies have begun shifting their focus from post-hoc explanations to inherently interpretable models. González-Abad et al. [14] demonstrate how XAI can be used to improve deep learning-based statistical downscaling processes by analyzing the internal behavior of models to improve spatial and temporal generalization [14]. Matera et al. [15] emphasize that a model's ability to explain the basis for its decisions is an essential requirement for the application of AI in predicting extreme climate events such as heat waves and floods, as model transparency is crucial for policymakers' confidence [15]. Furthermore, Mamalakis et al. [16] identify limitations of several XAI methods on convolutional neural networks (CNN) in geoscience, such as gradient shattering, which can reduce the quality of interpretation [16].

One increasingly emphasized innovative direction is the integration of scientific knowledge and local or indigenous

knowledge to enrich XAI-based climate prediction systems. According to Sohad and Mafrolla [17], local knowledge plays a crucial role in understanding the dynamics of environmental change at the micro level because it is rooted in the direct experiences of communities [17]. However, the literature still rarely combines the XAI framework with these traditional knowledge systems. A study by Okedele et al. [18] found that the integration of IKS into global climate change adaptation policies is essential for creating inclusive, equitable, and sustainable strategies [18]. IKS reflects ecological insights and sustainability practices developed by indigenous communities over centuries, offering unique perspectives on biodiversity conservation, natural resource management, and climate resilience. In a similar context, Bommer et al. [19] state that a systematic evaluation framework for assessing the robustness, faithfulness, and localization aspects of various XAI methods can serve as an interpretive bridge between data-driven models and scientific understanding of climate processes [19].

3. METHODOLOGY

This study uses a mixed-methods design that integrates qualitative ethnographic investigation with machine learning modeling to examine how IKS can be incorporated into interpretable climate predictions using XAI. The stages of this study are shown in Figure 1.

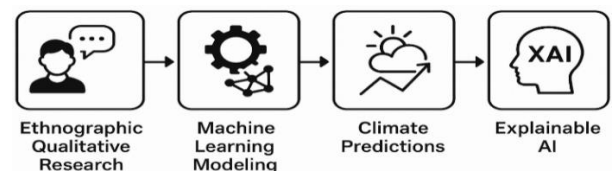


Figure 1. Research process flowchart

3.1 Ethnographic qualitative research

The first stage of this research consists of a qualitative ethnographic investigation aimed at systematically collecting and formalizing IKS held by local communities in North Maluku Province, Indonesia. Data were gathered through participatory observation, in-depth interviews, focused group discussions, and the documentation of traditional ecological practices related to natural signs and weather patterns.

Beyond descriptive documentation, this stage was designed to generate empirically grounded variables for computational modeling. Indigenous indicators—such as leaf curling, cloud formation over mountain peaks, seawater turbidity, insect activity, and bioluminescent plankton—were identified, coded, and transformed into structured categorical or binary representations. These representations serve as field-derived features reflecting lived ecological knowledge rather than hypothetical constructs.

The ethnographic data therefore constitute the primary empirical basis for integrating Indigenous Knowledge into the machine learning pipeline. Table 1 presents field-based ethnographic observations and their local interpretations that were systematically coded and transformed into input features for machine learning modeling. The table further demonstrates how these Indigenous Knowledge indicators were cross-validated using meteorological reference data from the Indonesian Meteorology, Climatology, and Geophysics Agency (BMKG).

Table 1. Indigenous knowledge indicators derived from ethnographic fieldwork for climate modeling

Data Source / Informant	Location / Indigenous Communities	Data Collection Methods	Observed Indicators / Signs of Nature	Local Meaning / Interpretation	Supporting Documentation	Additional Information
Indigenous leaders/elders of the community	Mare Gam, Tidore	In-depth interviews	Wind direction from the north, accompanied by the sound of certain insects	Signaling the start of the rainy season	Audio recordings, field notes	Data compared with rainfall predictions by the Indonesian Meteorology, Climatology, and Geophysics Agency
Traditional fishermen	Moti Island	Participatory observation	The color of the seawater became murkier, and the currents weakened	Signs of a small storm will occur in 1–2 days	Field photos and videos	Verified with wind speed data from weather sensors
Nutmeg and clove farmers	South Halmahera	Semi-structured interviews	Particular tree leaves curl up before it rains	Signs of high humidity	Field notes, results of qualitative coding	Used for feature extraction in Machine learning (ML) models
Local cultural figures	Ternate	Traditional documentation	Traditional song lyrics state 'clouds sleep on the peak of Gamalama'	Symbol of the arrival of the long dry season	Manuscripts, transcripts of indigenous texts	Contextualized with linguistic analysis
Coastal Communities	Maitara Island	Focused group discussion (FGD)	Changes in plankton bioluminescence patterns on the coast	Signaling sunny weather in the coming days	Night observation video	Potential converted into sensory visual variables

Table 1 demonstrates that the ethnographic dataset comprises not only interview transcripts but also visual observations, cultural symbols, environmental notes, and traditional artifacts derived from fieldwork. These elements were systematically coded and transformed into analytical variables that were directly incorporated as input features in the machine learning model, thereby enabling the explicit integration of Indigenous Knowledge with scientific climate data.

3.2 Ethical considerations and community engagement

This study involves IKS obtained through ethnographic fieldwork with local communities in North Maluku Province, Indonesia. Ethical considerations were integral to all stages of the research process. Participation in interviews, observations, and focus group discussions was entirely voluntary, and informed consent was obtained from all participants prior to data collection. The objectives of the study, the intended use of the collected knowledge, and the non-commercial nature of the research were clearly explained to community members. To protect participant privacy and community integrity, personal identifiers were anonymized, and sensitive cultural information was documented only with explicit permission.

Community engagement was conducted in a participatory manner, whereby indigenous indicators and their interpretations were discussed and validated collaboratively with community representatives to ensure that the recorded knowledge accurately reflected local meanings. This approach aimed to respect community agency and avoid misrepresentation of Indigenous Knowledge. The study does not claim ownership over Indigenous Knowledge and does not involve proprietary or commercial exploitation. Research findings are intended to support knowledge preservation and to inform transparent, explainable climate analysis frameworks that may benefit local communities through improved understanding of environmental dynamics.

3.3. Machine learning modeling

In this study, the model used to build a climate prediction system based on IKS data is Long Short-Term Memory

(LSTM), a variant of the Recurrent Neural Network (RNN) designed to process sequential or time-series data. LSTM was chosen based on its ability to recognize long-term temporal patterns and avoid the vanishing gradient problem that commonly occurs in conventional RNN models [20]. LSTM works by storing important information from past data sequences through a cell state and a gating system mechanism, which consists of three main gates: the input gate, the forget gate, and the output gate [21]. The input gate controls the new information to be stored, the forget gate determines the old information to be discarded, while the output gate regulates the output at each time step. With this mechanism, LSTM can effectively learn from sequential meteorological data such as daily rainfall, temperature, humidity, and wind direction, enabling accurate modeling of temporal dependencies in environmental and climatic systems [22–24].

LSTM was used to analyze the temporal relationships between environmental indicators and traditional knowledge. Data obtained from the ethnography stage (e.g., “curling leaves,” “clouds over mountain peak,” or “murky seawater”) were converted into numerical variables or binary categories. These data are then combined with quantitative climate variables such as temperature, rainfall, and humidity to form a multivariate time-series dataset. The LSTM model was trained with a training set comprising 80% of the total data, while 20% was used for testing (testing set). During training, the model learned patterns of environmental change that led to specific climate categories such as “sunny,” “cloudy,” or “rainy.” The LSTM prediction results are then evaluated using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and classification accuracy metrics to assess how well the model recognizes climate change patterns, combining scientific data with local knowledge.

This stage also lays the groundwork for the application of XAI in the next stage, where the LSTM model's output is explained using SHAP (SHapley Additive exPlanations) to determine which variables most influence the prediction results. The use of LSTM not only enables time-series-based learning but also enhances the transparency and interpretability of climate prediction results grounded in the Indigenous Knowledge System. Table 2 below presents the LSTM input data for climate prediction based on IKS.

Table 2. LSTM model input data for climate prediction based on Indigenous Knowledge Systems (IKS)

ID	Location	Time (Date)	Signs of Nature (IKS)	Rainfall (mm)	Temperature (°C)	Humidity (%)	Wind Direction (°)	Climate Label (Output)
MRG-001	Mare Gam, Tidore	2025-01-01	Leaves curled up	85	27.4	82	140	Rain
MRG-001	Mare Gam, Tidore	2025-01-02	Low, thick clouds	95	26.9	85	130	Rain
MRG-001	Mare Gam, Tidore	2025-01-03	Insects' activities at night	60	28.1	78	150	Cloudy
MTR-002	Maitara Island	2025-01-01	Bioluminescent plankton	5	30.5	70	210	Sunny
MTR-002	Maitara Island	2025-01-02	A weak wind from the west	10	31.0	68	220	Sunny

Note: LSTM = Long Short-Term Memory.

3.4 Climate predictions and Explainable AI

This stage continues the training of the LSTM model, using the developed model to predict climate conditions based on a combination of environmental indicators and traditional natural signs. These predictions are not only numerical (e.g., rainfall or temperature values) but also categorical, such as “sunny,” “cloudy,” or “rainy,” which are consistent with the terminology used by local indigenous communities.

In the climate prediction stage, the LSTM model receives input sequences of multivariate data comprising environmental parameters (temperature, humidity, rainfall, wind direction) and features derived from indigenous knowledge (e.g., leaf rolling, sea water color, or animal behavior). The model then generates output sequences as probabilistic predictions of the likelihood of each climate condition. These predictions can be displayed as temporal graphs, spatial maps, or climate category labels, depending on research needs. The next stage is to apply XAI to interpret how the model generates these predictions. XAI plays a vital role in providing transparency and accountability for artificial intelligence-based models, especially given the social and cultural dimensions of this research. Interpretation methods such as SHAP (SHapley Additive exPlanations) are used to explain each feature's contribution to the prediction. For example, if the LSTM model predicts “heavy rain,” the SHAP method can show that “high humidity” and “clouds covering mountain peaks” have the most significant influence on that prediction.

In addition, XAI analysis results can be visualized as heatmaps or feature importance plots to show which variables most strongly influence the prediction. This step ensures that the model results can be understood not only by researchers but also by indigenous peoples and policymakers, so that data-driven decisions continue to respect local knowledge and can be socially verified.

4. RESULTS AND DISCUSSION

4.1 Experimental results

This experiment tested the LSTM model's ability to predict climate conditions by integrating environmental variables and IKS indicators. The dataset used in this experiment was a field-informed multivariate time-series dataset representing ecological conditions across 12 locations in the North Maluku islands over 450 days of observation. The dataset integrates empirically derived Indigenous Knowledge indicators obtained from ethnographic fieldwork with observed

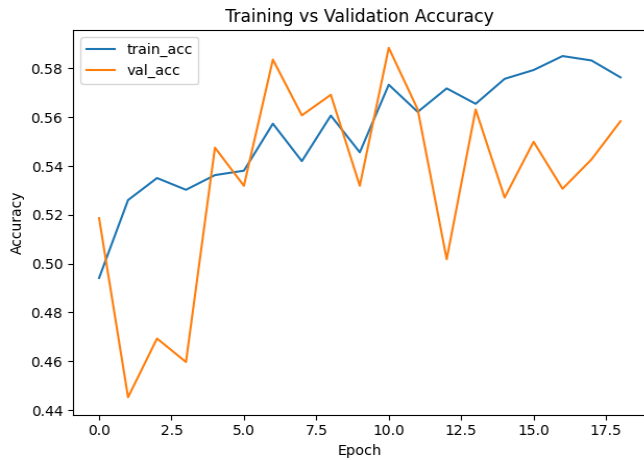
meteorological variables, with limited scenario-based augmentation applied solely to support experimental evaluation while preserving field-observed relationships. Each daily data point includes environmental variables such as rainfall, temperature, humidity, and wind direction, as well as IKS indicators such as leaf curling, cloud over mountains, muddy water, insect activity, and bioluminescent plankton. The data were normalized using StandardScaler and split into training (64%), validation (16%), and test (20%) sets. To capture temporal dynamics, the data were arranged in 14-day sequences (SEQ_LEN = 14) before being fed into the model. The model architecture consisted of a Bidirectional LSTM with 128 units followed by a 96-unit LSTM, dropout of 0.30 and 0.25 to prevent overfitting, and a Dense layer with ReLU and sigmoid activations on the output layer. The model was compiled with the Adam optimizer (learning rate 4×10^{-4}) and binary cross-entropy loss, then trained for up to 90 epochs with early stopping and ReduceLROnPlateau. After training, the model was evaluated on test data using accuracy, precision, recall, and F1-score metrics, and the results were visualized using training curves and confusion matrices. To interpret the results, the SHAP (SHapley Additive Explanations) method was applied to identify the relative contribution of each feature to the rainfall prediction. Figure 2 shows the training and validation curves.

Based on the training results shown in Figure 2, the LSTM model demonstrates a stable and focused learning process. In Figure 2(a), the training and validation accuracies gradually increase to around 0.58, indicating that the model can recognize data patterns well, even though the validation accuracy still shows fluctuations. Meanwhile, Figure 2(b) shows a consistent decrease in loss from around 0.76 to 0.68 without a significant increase in validation loss, indicating that overfitting did not occur. Overall, both curves suggest that the model has converged and maintains a balance between learning and adaptation to new data.

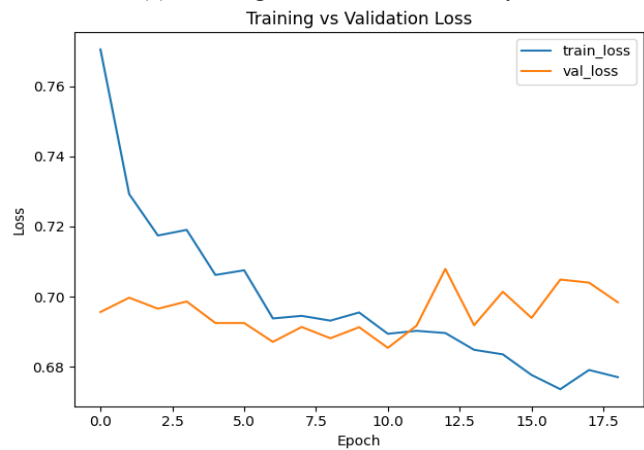
After the LSTM model converged during training, the next step was to evaluate its performance on previously unused test data. The evaluation assessed the model's ability to predict rainy and non-rainy conditions using a combination of environmental variables and IKS indicators. Figure 3 presents the model evaluation results as a confusion matrix.

Based on the training and evaluation results shown in Figure 3, the LSTM shows a classification bias towards the positive class (“Rainy”). Most of the actual rainfall data was correctly predicted by the model (high actual positive rate), reflected in the nearly perfect recall of 99.79. However, there are still many cases where the model predicts rain even when the exact condition is not rainy (false positives), resulting in a precision of 51.37%. As a result, the model's overall accuracy is only

51.32%, with an F1-score of 0.6783, indicating a moderate balance between rain detection and prediction accuracy.



(a) Training vs Validation accuracy



(b) Training vs Validation loss

Figure 2. Training and validation curves of the Long Short-Term Memory (LSTM) model

Furthermore, to interpret the LSTM model's decisions, SHAP (SHapley Additive exPlanations) analysis was performed on the test sample. Since the model input is a time series, each feature is evaluated based on its contribution at each time step, and these values are averaged over 2 weeks to obtain the overall importance of each feature. The results of this analysis are shown in Figure 4, which displays the distribution of each variable's contribution to the model's decision when predicting the probability of rain. The color of the dots represents the original value of the feature (red for high values, blue for low values). In contrast, the position of the dots along the horizontal axis indicates the direction and magnitude of their influence on the prediction results: dots that shift to the right increase the probability of rain (Rainy), while those that change to the left decrease the probability (Non-Rainy).

Based on the SHAP analysis results in Figure 4, environmental features such as rainfall, humidity, and temperature have the highest contribution values to the model prediction, indicating that these three variables are the main determinants influencing the probability of rainfall. This is consistent with the characteristics of a humid tropical climate in archipelagic regions such as North Maluku, where increased humidity and decreased temperature are often early indicators of convective cloud formation. In addition, local knowledge

indicators such as clouds over mountains and leaf curling also make a significant positive contribution to rainfall prediction. These two indicators represent ecological signals that indigenous peoples have long used to indicate weather change. In contrast, the bioluminescent plankton feature shows a negative association with the probability of rain, consistent with the calm seas and clear skies that often accompany this phenomenon. Overall, these XAI results indicate that Indigenous Knowledge indicators contribute to the model's predictions alongside meteorological variables. However, feature contribution alone does not constitute definitive evidence of meaningful knowledge integration, which requires alignment between the learned contribution patterns and the underlying ecological interpretations within IKS.

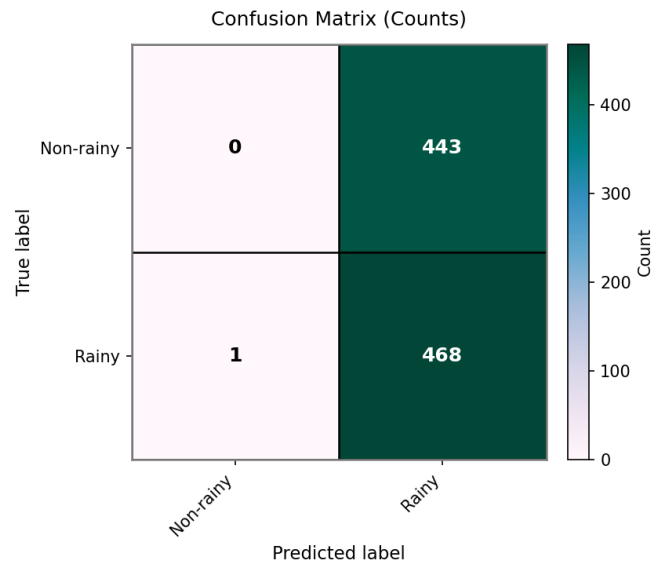


Figure 3. Confusion matrix of model evaluation result

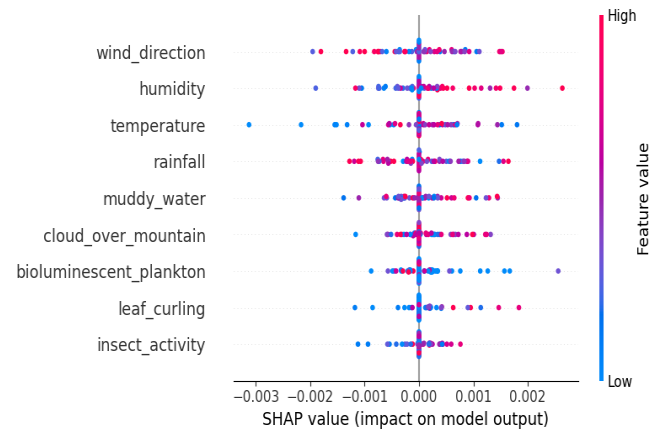


Figure 4. SHapley Additive exPlanations (SHAP) analysis results

To assess whether the observed SHAP contributions reflect meaningful integration rather than statistical coincidence, the contribution patterns of selected Indigenous Knowledge indicators were examined in relation to their ecological interpretations. For instance, the “clouds over mountains” indicator exhibits positive SHAP values predominantly under conditions of increased humidity and rainfall probability, which is consistent with its traditional interpretation as an early signal of the rainy season. Similarly, the positive

contribution of the leaf curling indicator aligns with pre-rainfall moisture accumulation observed in local ecological practices.

Nevertheless, this analysis remains indicative rather than conclusive. While the observed alignment suggests semantic consistency between model behavior and Indigenous Knowledge interpretations, the study does not perform formal causal validation across multiple seasonal cycles. As such, the results should be interpreted as evidence of preliminary alignment rather than definitive proof of full knowledge integration. Further work is required to evaluate the robustness of these patterns across extended temporal and spatial contexts.

4.2 Root-cause analysis of model performance

Despite the successful convergence of the LSTM model and its ability to capture temporal patterns, the overall classification accuracy remains moderate. To address this limitation, a root-cause analysis was conducted to examine the underlying factors contributing to the observed performance, particularly the imbalance between recall and precision.

First, data distribution and classification threshold effects play a central role. The dataset exhibits a dominance of rainfall-related events, which biases the learning process toward the positive (“Rainy”) class. This imbalance, combined with a relatively low decision threshold, results in very high recall (99.79%) but increased false positives, thereby reducing precision and overall accuracy. While this behavior is suboptimal for accuracy-oriented evaluation, it is consistent with the model’s design objective as a risk-averse early warning system, where minimizing missed rainfall events is prioritized over avoiding false alarms.

Second, the encoding strategy of IKS indicators introduces additional limitations. In the current implementation, several indigenous indicators—such as leaf curling, cloud formation over mountain peaks, and seawater turbidity—are represented using binary or coarse categorical encodings. Although this approach enables straightforward integration into the machine learning pipeline, it inevitably simplifies complex ecological phenomena. Important nuances, such as the intensity, duration, or gradual progression of these indicators, may be lost during discretization. This reduction in representational richness can weaken the discriminative power of IKS features and partially explain the model’s limited accuracy.

Third, the suitability of the LSTM architecture for hybrid feature sets must be considered. LSTM networks are well suited for modeling continuous numerical time-series data, such as rainfall, temperature, and humidity. However, the hybrid nature of the input—combining sequential meteorological variables with categorical or semi-static indigenous indicators—may not be optimally handled by a purely recurrent architecture. The lack of dedicated embedding layers or attention mechanisms for categorical features may constrain the model’s ability to fully exploit the informational content of IKS indicators.

Finally, dataset construction and augmentation constraints may also contribute to performance limitations. Although the dataset is field-informed and grounded in ethnographic observations, the temporal and spatial coverage remains limited. Scenario-based augmentation was applied only to support experimental evaluation, but such augmentation may still introduce simplified patterns that do not fully reflect the causal complexity of real-world climate–ecology interactions.

This limitation can lead to spurious correlations and reduce generalization performance on unseen data.

Overall, this root-cause analysis indicates that the moderate accuracy observed in the model is not attributable to a single factor, but rather to the combined effects of data imbalance, simplified feature encoding, architectural constraints, and dataset scale. These findings underscore the importance of aligning evaluation metrics with application goals and motivate future improvements, including richer encoding schemes for indigenous indicators, hybrid or attention-based model architectures, and the expansion of field-validated joint datasets.

4.3 Discussion

The experimental results demonstrate that the LSTM-based sequence learning approach is capable of capturing temporal relationships between meteorological variables and IKS indicators. The high recall value indicates that the model is highly sensitive to rainfall-related signals, reflecting its tendency to prioritize the detection of rainy conditions based on both scientific climate variables (e.g., rainfall and humidity) and selected indigenous ecological indicators such as clouds over mountains and leaf curling. This behavior is consistent with a risk-averse early warning perspective, where minimizing missed rainfall events is prioritized.

However, the relatively low precision and overall accuracy reveal important limitations that must be interpreted carefully. The observed bias toward the positive class results in frequent false alarms, which can be attributed to data imbalance, simplified encoding of IKS indicators, and architectural constraints of the LSTM when handling hybrid numerical–categorical feature sets. These limitations indicate that the proposed model should not be interpreted as an optimized or operational climate prediction tool.

Importantly, the primary contribution of this study does not lie in achieving high predictive accuracy, but in proposing and empirically examining a methodological framework for integrating IKS with scientific climate data using XAI. Within this framework, XAI is employed not to improve raw model performance, but to provide interpretability and transparency regarding how indigenous indicators and meteorological variables jointly influence model decisions. This distinction is crucial, as SHAP-based explanations reveal patterns of feature contribution that can be compared against established indigenous ecological interpretations, thereby enabling an initial assessment of semantic alignment rather than definitive causal validation.

From this perspective, the results should be understood as a proof-of-concept and exploratory validation of knowledge integration, demonstrating how indigenous ecological indicators can be operationalized, encoded, and interpreted within a data-driven modeling pipeline. While the current implementation exhibits performance limitations, it provides a transparent foundation upon which more advanced architectures, richer encoding schemes, and larger field-validated datasets can be developed.

Future research should therefore focus on improving representational fidelity of IKS indicators (e.g., ordinal or fuzzy encoding), exploring hybrid or attention-based models to better handle heterogeneous feature types, and conducting longitudinal validation across multiple seasonal cycles. These steps are necessary before the framework can be extended toward robust, application-oriented climate services.

5. CONCLUSION

This study explored the integration of scientific climate variables and IKS within an XAI framework to examine how local ecological indicators can be operationalized and interpreted in data-driven climate analysis. Using an LSTM-based time-series model, the study demonstrated that short-term temporal patterns between meteorological factors and selected indigenous indicators can be learned and analyzed in an interpretable manner.

Although the overall predictive accuracy of the model was moderate (0.5132), the high recall (99.79%) highlights the model's strong sensitivity to rainfall-related signals, while the lower precision (51.37%) reflects a tendency toward overprediction. These results indicate that the current implementation should not be interpreted as an optimized climate prediction system, but rather as an exploratory and proof-of-concept framework. SHAP-based explanations revealed that meteorological variables such as rainfall, humidity, and temperature dominate model decisions, while Indigenous Knowledge indicators—including clouds over mountain peaks and leaf curling—exhibit consistent contributions that are broadly aligned with their ecological interpretations.

The primary contribution of this study lies in proposing and empirically examining a transparent methodological framework for integrating Indigenous Knowledge and scientific data using XAI, rather than in achieving high predictive performance. By emphasizing interpretability and semantic alignment, this work provides an initial foundation for ethically and transparently incorporating indigenous ecological knowledge into climate-related artificial intelligence research.

Future research should focus on expanding field-validated datasets, improving the representational fidelity of Indigenous Knowledge indicators through richer encoding schemes, and exploring hybrid or attention-based model architectures to better handle heterogeneous feature types. These efforts are necessary to move from exploratory analysis toward more robust, context-aware climate services that remain transparent, interpretable, and respectful of community knowledge systems.

REFERENCES

- [1] Bawack, R., Roderick, S., Badhrus, A., Dennehy, D., Corbett, J. (2025). Indigenous knowledge and information technology for sustainable development. *Information Technology for Development*, 31(2): 233-250. <https://doi.org/10.1080/02681102.2025.2472495>
- [2] Sharma, A., Sharma, D., Grewal, A.S., Bajaj, H., Yadav, M., Dhingra, A.K., Chopra, B. (2024). Chapter 8- Importance of indigenous knowledge in achieving environmental sustainability. In *Role of Green Chemistry in Ecosystem Restoration to Achieve Environmental Sustainability*, pp. 75-82. <https://doi.org/10.1016/B978-0-443-15291-7.00015-8>
- [3] Tharakan, J. (2017). Indigenous knowledge systems for appropriate technology development. *Indigenous People*, 123(1): 123-134. <https://doi.org/10.5772/intechopen.69889>
- [4] Martin, J.F., Roy, E.D., Diemont, S.A., Ferguson, B.G. (2010). Traditional Ecological Knowledge (TEK): Ideas, inspiration, and designs for ecological engineering. *Ecological Engineering*, 36(7): 839-849. <https://doi.org/10.1016/j.ecoleng.2010.04.001>
- [5] Mbah, M., Ajaps, S., Molthan-Hill, P. (2021). A systematic review of the deployment of indigenous knowledge systems towards climate change adaptation in developing world contexts: Implications for climate change education. *Sustainability*, 13(9): 4811. <https://doi.org/10.3390/su13094811>
- [6] Motsumi, M.M., NemaKonde, L.D. (2025). Indigenous early warning indicators for improving natural hazard predictions. *Jambá: Journal of Disaster Risk Studies*, 17(1): 12. <https://doi.org/10.4102/jamba.v17i1.1754>
- [7] Jiri, O., Mafongoya, P., Chivenge, P. (2015). Indigenous knowledge systems, seasonal 'quality' and climate change adaptation in Zimbabwe. *Climate Research*, 66(2): 103-111. <https://doi.org/10.3354/cr01334>
- [8] Mafongoya, O., Mafongoya, P.L., Mudhara, M. (2021). Using indigenous knowledge systems in seasonal prediction and adapting to climate change impacts in bikita district in Zimbabwe. *The Oriental Anthropologist: A Bi-annual International Journal of the Science of Man*, 21(1): 195-209. <https://doi.org/10.1177/0972558x21997662>
- [9] Ali, S., Abuhmed, T., El-Sappagh, S., Muhammad, K., Alonso-Moral, J.M., Confalonieri, R., Guidotti, R., Del Ser, J., Díaz-Rodríguez, N., Herrera, F. (2023). Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. *Information Fusion*, 99: 101805. <https://doi.org/10.1016/j.inffus.2023.101805>
- [10] Saarela, M., Podgorelec, V. (2024). Recent applications of explainable AI (XAI): A systematic literature review. *Applied Sciences*, 14(19): 8884. <https://doi.org/10.3390/app14198884>
- [11] Muriithi, D.K., Lumumba, V.W., Awe, O.O., Muriithi, D.M. (2025). An explainable artificial intelligence models for predicting malaria risk in Kenya. *European Journal of Artificial Intelligence and Machine Learning*, 4(1): 1-8.
- [12] Huang, F., Jiang, S., Li, L., Zhang, Y., Zhang, Y., Zhang, R., Dai, Y. (2024). Applications of Explainable artificial intelligence in Earth system science. *arXiv preprint arXiv:2406.11882*. <https://doi.org/10.48550/arXiv.2406.11882>
- [13] Schiller, J., Stiller, S., Ryo, M. (2025). Artificial intelligence in environmental and Earth system sciences: Explainability and trustworthiness. *Artificial Intelligence Review*, 58(10): 1-23. <https://doi.org/10.1007/s10462-025-11165-2>
- [14] González-Abad, J., Baño-Medina, J., Gutiérrez, J.M. (2023). Using explainability to inform statistical downscaling based on deep learning beyond standard validation approaches. *Journal of Advances in Modeling Earth Systems*, 15(11): e2023MS003641. <https://doi.org/10.1029/2023ms003641>
- [15] Materia, S., García, L.P., van Straaten, C., O, S., Mamalakis, A., Cavicchia, L., Coumou, D., de Luca, P., Kretschmer, M., Donat, M. (2024). Artificial intelligence for climate prediction of extremes: State of the art, challenges, and future perspectives. *Wires Climate Change*, 15(6): e914. <https://doi.org/10.1002/wcc.914>
- [16] Mamalakis, A., Barnes, E.A., Ebert-Uphoff, I. (2022). Investigating the fidelity of explainable artificial

- intelligence methods for applications of convolutional neural networks in geoscience. *Artificial Intelligence for the Earth Systems*, 1(4): e220012. <https://doi.org/10.1175/aies-d-22-0012.1>
- [17] Sohad, M.K.N., Mafrolla, E. (2025). Bridging science and society: The integration of indigenous and scientific knowledge management. *Journal of Knowledge Management*, 29(7): 2258-2284. <https://doi.org/10.1108/jkm-11-2024-1326>
- [18] Okedele, P.O., Aziza, O.R., Oduro, P., Ishola, A.O. (2024). Integrating indigenous knowledge systems into global climate adaptation policies. *International Journal of Engineering Research and Development*, 20(12): 223-231.
- [19] Bommer, P.L., Kretschmer, M., Hedström, A., Bareeva, D., Höhne, M.M. (2024). Finding the right XAI method- A guide for the evaluation and ranking of explainable AI methods in climate science. *Artificial Intelligence for the Earth Systems*, 3(3): e230074. <https://doi.org/10.1175/aies-d-23-0074.1>
- [20] Pinjarkar, L., Sagayamary, S., P, R., Lebaka, S., Srinivas, P., Sandiri, R., Ramasamy, J., C, S. (2025). Prediction of respiratory tract infections using IoT and RNN techniques. *Engineering, Technology & Applied Science Research*, 15(5): 27250-27256. <https://doi.org/10.48084/etasr.11642>
- [21] Malashin, I., Tynchenko, V., Gantimurov, A., Nelyub, V., Borodulin, A. (2024). Applications of Long Short-Term Memory (LSTM) networks in polymeric sciences: A review. *Polymers*, 16(18): 2607. <https://doi.org/10.3390/polym16182607>
- [22] Waqas, M., Humphries, U.W. (2024). A critical review of RNN and LSTM variants in hydrological time series predictions. *MethodsX*, 13: 102946. <https://doi.org/10.1016/j.mex.2024.102946>
- [23] Krichen, M., Mihoub, A. (2025). Long short-term memory networks: A comprehensive survey. *AI*, 6(9): 215. <https://doi.org/10.3390/ai6090215>
- [24] Trstenjak, B., Brekalo, S., Trstenjak, J. (2025). Air quality index forecasting in the HRcity smart city system based on an LSTM prediction model. *Engineering, Technology & Applied Science Research*, 15(4): 24820-24824. <https://doi.org/10.48084/etasr.11665>