



## Temporally Validated Machine Learning for Solar Radiation as a Foundation for UV-Risk Modeling in High-Altitude Puno

Leonel Coyla Idme<sup>1\*</sup>, Edgar Eloy Carpio Vargas<sup>1</sup>, José Pánfilo Tito Lipa<sup>1</sup>, Ernesto Nayer Tumi Figueroa<sup>1</sup>

<sup>1</sup>*School of Statistical and Computer Engineering, Universidad Nacional del Altiplano, Puno, Peru.*

### ABSTRACT

High-altitude UV-risk services require predictive models whose target, validation design, and atmospheric predictors are physically aligned. This study establishes a reproducible machine-learning foundation for that objective by modeling the quantity actually measured at the Puno automatic station: hourly global solar radiation. We analyzed 43,739 records from 2020–2024 and formulated prediction as supervised regression under a prospective temporal holdout. Calendar and wind direction were mapped to cyclic coordinates; predictor gaps were imputed using training-only medians; and 2020–2023 observations were used to estimate a month-hour climatology, ordinary least squares, ridge regression, and gradient-boosted decision stumps. Performance was evaluated exclusively on 2024 using MAE, RMSE, and  $R^2$ . The climatology yielded the best full-day generalization (MAE 57.6 W/m<sup>2</sup>, RMSE 116.9 W/m<sup>2</sup>,  $R^2$  0.896), outperforming boosting in RMSE by 3.0%. Boosting reduced daytime RMSE by only 1.0% but increased daytime MAE by 10.8%. These results identify deterministic solar geometry as the dominant predictable component and missing optical variables as the principal barrier to short-term accuracy. The framework is therefore a validated benchmark for future UV-risk modeling, not a surrogate UV-index predictor; operational extension requires colocated erythemal UV, cloud, ozone, aerosol, and albedo measurements.

**Keywords:** Solar radiation, Temporal validation, Machine learning, High altitude, Puno

**Corresponding author:** Leonel Coyla Idme

**e-mail** ✉ [lcoyla@unap.edu.pe](mailto:lcoyla@unap.edu.pe)

**Received:** 02 August 2026

**Accepted:** 08 September 2026

### INTRODUCTION

High-altitude communities need dependable radiation information because exposure varies strongly within each day and across seasons. Ultraviolet (UV) and solar-radiation forecasting has therefore attracted statistical and machine-learning approaches, including seasonal models, neural networks, random forests, and recurrent models (Junk *et al.*, 2007; Oliveira *et al.*, 2021; Dabeedol *et al.*, 2023; Jiang *et al.*, 2024; Al-Farsi *et al.*, 2026). These studies also show that useful UV models require a measured UV target or a physically relevant satellite product.

The Puno project originally sought to predict the UV index from local meteorology. Inspection of the supplied station data changed the estimand: the response is hourly global solar radiation in W/m<sup>2</sup>, not erythemally weighted UV radiation or UV index. Treating these quantities as interchangeable would produce a precise-looking but invalid health indicator. This study therefore asks a narrower, answerable question: how well can global solar radiation be predicted from the available hourly station variables, and how much does a nonlinear model improve on a strong temporal baseline?

This work contributes an auditable five-year analysis rather than a claim of UV prediction. First, it uses a prospective 2024 test period to preserve temporal ordering. Second, it compares

meteorological regressors against a month-hour climatology, making the dominant periodic signal explicit. Third, it releases the complete processing code, tables, figures, predictions, and checksums. The central finding is that complexity does not guarantee better full-day prediction: the temporal baseline explains 89.6% of test variance.

### Three questions turn the project idea into testable decisions

The study is organized around three decisions that a future monitoring system must make. The first question is whether the supplied variables support a valid UV-index model; the answer depends on whether UV is measured. The second question is whether contemporaneous meteorology improves the prediction of the measured outcome, global solar radiation, beyond its recurring calendar pattern. The third question is whether any apparent gain survives a chronological test that represents a future year. Stating these decisions before interpreting the models prevents the analysis from drifting toward whichever result appears most favorable.

Each question has a corresponding evidentiary threshold. A UV claim requires a UV target and therefore cannot be tested with this workbook. A meteorological model must outperform the month-hour climatology on the untouched 2024 period to justify its added complexity. Finally, performance must be interpreted with both full-day and positive-radiation metrics because nighttime zeros can make a model appear more accurate than it is during exposure-relevant hours. These thresholds transform model comparison from a ranking exercise into a defensible decision rule.

## Related work

*Solar and ultraviolet radiation forecasting*

The closest literature separates the prediction target from its atmospheric controls. Junk *et al.* (2007) reconstructed UV irradiation with radiation, visibility, and ozone-related information, whereas Jiang *et al.* (2024) combined ground UV observations with satellite predictors and evaluated several cross-validation schemes. Oliveira *et al.* (2021) treated the UV index as a multistep sequence problem, and Dabeedool *et al.* (2023) compared neural and seasonal time-series models for both monthly solar radiation and UV index. Adjacent work extends sequence learning to long-horizon solar activity (Marcucci *et al.*, 2023), total-column ozone (Wright & Matthews, 2025; Hashemzadeh *et al.*, 2026), and the calibration stability of reflective solar sensors (Mohammad *et al.*, 2024; T.-C. Liu *et al.*, 2025). Aman *et al.* (2025) further show that cloud cover and UV index can improve photovoltaic prediction when those variables are measured. Together, these studies support two requirements for the present analysis: the target must be defined exactly, and complex models must be compared with seasonal baselines.

*Meteorological machine learning across environmental systems*

A broader environmental literature demonstrates the value—and the limitations—of nonlinear meteorological models. XGBoost, random forests, mixed spatial models, and attribution frameworks have been used to study ozone and particulate matter under changing weather regimes (Hsu *et al.*, 2024; Li *et al.*, 2025; F. Liu *et al.*, 2025; S. Liu *et al.*, 2025; Meyer *et al.*, 2025; Tong *et al.*, 2025; Wang *et al.*, 2025; Wan *et al.*, 2026). Integrated forecasting systems consequently emphasize cyclic variables, transparent baselines, residual diagnostics, and interpretable contributions (Ashraf *et al.*, 2025; Barkani & Allouani, 2025). Forecasts have also been embedded in adaptive urban infrastructure and edge-computing systems (Bedi *et al.*, 2026; Bouali *et al.*, 2026). These applications motivate the present feature engineering, but they do not remove the need for variables that directly represent clouds and atmospheric optical conditions.

*Health, exposure, and warning applications*

Environmental prediction becomes useful only when the exposure pathway is explicit. Meteorological machine learning has supported dengue forecasting in Taiwan and Singapore (Kuo *et al.*, 2024; Tian *et al.*, 2024), while climate and UV indicators have been examined in models of COVID-19 and influenza activity (Ianevski *et al.*, 2019; Scabbia *et al.*, 2022; Lyu *et al.*, 2023; Chandrasekhar *et al.*, 2025). More direct radiation applications include sunburn-risk prediction for outdoor workers (Gong *et al.*, 2022), automated weather-alert frameworks (Bainapalepu & Bathula, 2025), and image-based characterization of skin lesions (Sanchez *et al.*, 2021). These studies establish the social relevance of radiation information, but they also caution against translating a global radiation estimate into a health index without a validated dose-response or UV calibration.

*High-altitude context and transferability*

Puno's altitude makes transferability a scientific issue rather than a geographic detail. High-altitude research documents UV-related aging outcomes, ecosystem responses, biological adaptation, and distinctive microbial environments (Lei *et al.*, 2022; Chen *et al.*, 2023; Zhu *et al.*, 2023; Collins *et al.*, 2025; L. Wu *et al.*, 2025; Teklu *et al.*, 2025). Geography and climate also leave detectable signatures in agricultural products (Y. Wu *et al.*, 2026), while electromagnetic snow sensing illustrates the instrumentation challenges of cold, elevated environments (Alexander & Patel, 2025). UV exposure affects engineered materials as well as organisms (Bi *et al.*, 2026), and outdoor-productivity research shows that environmental effects can be nonlinear and occupation-specific (X. Liu *et al.*, 2018; Brown & Martinez, 2024). This diversity justifies local validation: a model trained elsewhere cannot be assumed to retain its calibration in the Altiplano.

## MATERIALS AND METHODS

*One station provides five years of hourly evidence*

The supplied workbook contains hourly observations from the Puno automatic meteorological station from 1 January 2020 through 31 December 2024. Variables include atmospheric pressure, air temperature, relative humidity, accumulated precipitation, wind speed, wind direction, and global solar radiation. The two supplied workbooks have the same SHA-256 hash and are exact duplicates; one copy was retained in the reproducibility package. Global solar radiation ( $\text{W}/\text{m}^2$ ) was the outcome.

*Cleaning preserves the raw record and isolates auditable transformations*

Raw measurements were not overwritten. We parsed date and hour into a timestamp, sorted records chronologically, and removed only the 230 rows without radiation measurements from modeling. Missing predictors were imputed with medians calculated on 2020–2023 data. Hour, day of year, month, and wind direction were represented by sine and cosine pairs to preserve their circular geometry. Negative radiation values were absent.

*The prediction problem is a temporally indexed regression task*

Let  $y_t$  denote global solar radiation in  $\text{W}/\text{m}^2$  at hour  $t$ , and let  $\mathbf{x}_t$  contain pressure, temperature, relative humidity, accumulated precipitation, wind speed, wind direction, and calendar variables. The learning objective is to estimate a nonnegative mapping  $f_\theta$  from the information available at  $t$  to the conditional expectation of radiation. This formulation makes the estimand explicit and prevents the UV index from entering the model as an unobserved proxy.

$$\hat{y}_t = \max\{0, f_\theta(\mathbf{x}_t)\} \approx \mathbb{E}[y_t | \mathbf{x}_t] \quad (1)$$

Periodic variables were embedded on the unit circle to eliminate artificial discontinuities between 23:00 and 00:00, December and January, or  $359^\circ$  and  $0^\circ$ . For a periodic variable  $z$  with period  $P$ , the transformation is

$$\phi(z; P) = [\sin(2\pi z/P), \cos(2\pi z/P)]^T \quad (2)$$

The feature vector concatenates six continuous meteorological variables with sine-cosine pairs for hour, day of year, month, and wind direction as you see in **Table 1**. Missing continuous values are replaced by the median  $m_j$  estimated from the training period only, and standardized values use the training mean  $\mu_j$  and standard deviation  $\sigma_j$ .

$$\tilde{x}_{tj} = (x_{tj} - \mu_j)/\sigma_j, \quad x_{tj} \leftarrow m_j \text{ when } x_{tj} \text{ is missing} \quad (3)$$

**Table 1.** Coverage and quality of the supplied hourly variables.

| Variable                            | Valid n | Missing % | Median | Maximum |
|-------------------------------------|---------|-----------|--------|---------|
| Pressure (mmb)                      | 43560   | 0.4       | 646.8  | 746.0   |
| Temperature (°C)                    | 43655   | 0.2       | 9.9    | 21.8    |
| Humidity (%)                        | 41484   | 5.2       | 50.0   | 100.0   |
| Precipitation (mm)                  | 42751   | 2.3       | 0.0    | 234.0   |
| Wind speed (m/s)                    | 43586   | 0.3       | 2.1    | 21.2    |
| Wind direction (°)                  | 43581   | 0.4       | 173.0  | 360.0   |
| Solar radiation (W/m <sup>2</sup> ) | 43509   | 0.5       | 7.0    | 1287.0  |

#### A future-year test separates prediction from interpolation

All models used 34,788 observations from 2020–2023 for training and 8,721 observations from 2024 for testing. The baseline predicts each test hour with the training mean for its month-hour combination. Linear and ridge regression represent additive relationships. Gradient boosting sequentially combines 80 depth-one decision trees with a learning rate of 0.08, allowing nonlinear threshold effects while remaining inspectable. Predictions were constrained to nonnegative values.

#### Four estimators isolate seasonality, shrinkage, and nonlinearity

The climatological estimator conditions only on calendar month  $m$  and hour  $h$ . It is intentionally strong because solar geometry is highly periodic; any higher-capacity model must demonstrate incremental information beyond this benchmark.

$$\hat{y}_{t, \text{clim}}^{\text{clim}} = (1/N_{mh}) \sum_i \in \text{Train } y_i \cdot \mathbb{1}(\text{month}_i = m, \text{hour}_i = h) \quad (4)$$

Ordinary least squares estimates a global additive response, whereas ridge regression stabilizes correlated cyclic and meteorological coefficients through  $L_2$  regularization. The intercept is not penalized.

$$\hat{\beta}_{\text{ridge}} = \arg \min_{\beta} \{ \sum_i \in \text{Train } (y_i - \beta_0 - \tilde{x}_i^T \beta)^2 + \lambda \|\beta\|_2^2 \} \quad (5)$$

The nonlinear estimator is an additive ensemble of  $M=80$  depth-one regression trees. At boosting iteration  $m$ , a stump  $h_m$  partitions one standardized predictor at a learned threshold and approximates the current residual;  $v=0.08$  controls each update.

$$\begin{aligned} F_0(x) &= \bar{y}_{\text{train}}, \quad F_m(x) \\ &= F_{m-1}(x) + v h_m(x), \quad m \\ &= 1, \dots, 80 \end{aligned} \quad (6)$$

#### Complementary metrics separate typical error from large misses

Performance was summarized by mean absolute error (MAE), root mean square error (RMSE), and the coefficient of determination ( $R^2$ ). Metrics were computed for all observations and separately for positive-radiation hours. This second analysis prevents the many nighttime zeros from obscuring daytime error. The full procedure is deterministic and uses no future target values.

$$\begin{aligned} \text{MAE} &= (1/n) \sum_{i=1}^n |y_i - \hat{y}_i|, \quad \text{RMSE} \\ &= \sqrt{[(1/n) \sum_{i=1}^n (y_i - \hat{y}_i)^2]} \end{aligned} \quad (7)$$

$$R^2 = 1 - [\sum_{i=1}^n (y_i - \hat{y}_i)^2 / \sum_{i=1}^n (y_i - \bar{y})^2] \quad (8)$$

MAE estimates the typical absolute deviation, RMSE emphasizes infrequent large misses, and  $R^2$  measures explained variation relative to the 2024 test mean. Reporting all three prevents a single metric from concealing an operationally relevant failure mode.

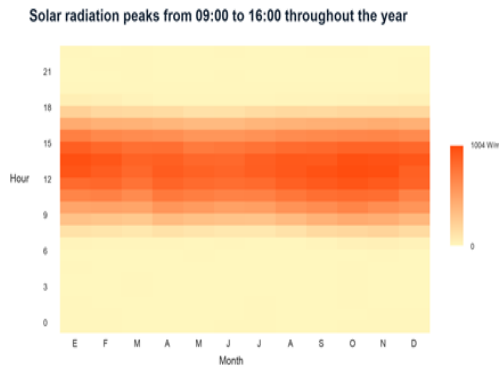
## RESULTS AND DISCUSSION

#### The record is nearly complete and strongly periodic

The target is available for 99.5% of rows, while relative humidity has the largest predictor gap (5.2%); (**Table 1**). Mean monthly radiation ranges from 247.0 W/m<sup>2</sup> in June to 304.2 W/m<sup>2</sup> in October (**Table 2**). The median remains close to zero because each month includes nighttime observations. Radiation is concentrated between approximately 09:00 and 16:00 throughout the year (**Figure 1**), indicating that any credible model must first match the solar cycle.

**Table 2.** Monthly climatology of global solar radiation.

| Month | n    | Mean W/m <sup>2</sup> | P95 W/m <sup>2</sup> | Positive % |
|-------|------|-----------------------|----------------------|------------|
| Jan   | 3685 | 280.1                 | 1078                 | 56.1       |
| Feb   | 3365 | 270.1                 | 1060                 | 55.1       |
| Mar   | 3700 | 250.0                 | 991                  | 53.6       |
| Apr   | 3593 | 267.6                 | 975                  | 53.4       |
| May   | 3698 | 252.2                 | 898                  | 51.4       |
| Jun   | 3576 | 247.0                 | 866                  | 51.3       |
| Jul   | 3680 | 256.4                 | 895                  | 51.1       |
| Aug   | 3695 | 276.1                 | 970                  | 52.4       |
| Sep   | 3580 | 289.2                 | 1050                 | 53.7       |
| Oct   | 3689 | 304.2                 | 1108                 | 54.1       |
| Nov   | 3553 | 301.6                 | 1095                 | 54.4       |
| Dec   | 3695 | 277.2                 | 1041                 | 55.2       |



**Figure 1.** Mean hourly solar radiation by month. The dominant structure is diurnal; seasonal changes mainly affect intensity.

#### *A simple temporal baseline is difficult to beat*

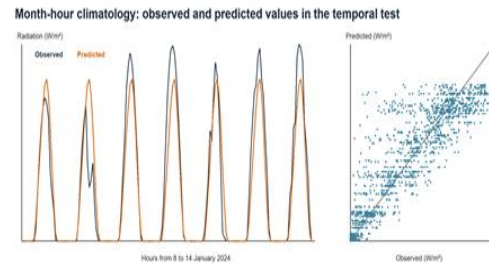
The month-hour climatology produced the lowest full-day error: MAE 57.6 W/m<sup>2</sup>, RMSE 116.9 W/m<sup>2</sup>, and R<sup>2</sup> 0.896 (**Table 3**). Boosting was close but did not improve full-day RMSE. Linear and ridge regression performed substantially worse, indicating that additive effects cannot represent the sharp transition between nighttime and daytime even with cyclic calendar features.

**Table 3.** Prospective performance on the 2024 test set; lower MAE/RMSE and higher R<sup>2</sup> are better.

| Model                   | Hours    | MAE   | RMSE  | R <sup>2</sup> |
|-------------------------|----------|-------|-------|----------------|
| Month-hour climatology  | All      | 57.6  | 116.9 | 0.896          |
| Month-hour climatology  | Positive | 107.2 | 159.7 | 0.800          |
| Linear regression       | All      | 125.3 | 168.6 | 0.783          |
| Linear regression       | Positive | 175.6 | 209.3 | 0.657          |
| Ridge regression        | All      | 125.3 | 168.5 | 0.783          |
| Ridge regression        | Positive | 175.6 | 209.3 | 0.657          |
| Decision-stump boosting | All      | 73.3  | 120.6 | 0.889          |
| Decision-stump boosting | Positive | 118.7 | 158.1 | 0.804          |

#### *Daytime variability exposes the missing information*

During positive-radiation hours, boosting achieved a slightly lower RMSE than climatology (158.1 versus 159.7 W/m<sup>2</sup>), but climatology retained the lower MAE (107.2 versus 118.7 W/m<sup>2</sup>). **Figure 2** shows why: the baseline tracks the regular daily envelope but cannot reproduce abrupt attenuation or unusually high peaks. Rapid cloud changes are a known source of short-term radiation variability (Junk *et al.*, 2007), yet cloud cover, total-column ozone, aerosols, and direct UV measurements are absent here. The marginal daytime gain from boosting is therefore scientifically plausible but insufficient for an operational UV warning system.



**Figure 2.** The best full-day model captures the daily envelope but smooths short-lived departures in the 2024 test period.

#### *Metric-dependent rankings reveal a small, not decisive, nonlinear gain*

The model ranking changes only when the evaluation emphasizes large daytime misses. Across all hours, climatology reduces RMSE by 3.0% relative to boosting and reduces MAE by 21.4%. During positive-radiation hours, boosting reduces RMSE by only 1.0%, while climatology still reduces MAE by 9.7%. Thus, boosting occasionally follows daytime peaks more closely, but it makes larger typical errors. This metric-dependent trade-off is too small to justify replacing the transparent baseline for general hourly estimation, although it identifies the specific regime in which better optical predictors could add value.

#### *The error structure identifies the missing physical state*

The residual pattern is consistent with an omitted-variable mechanism rather than insufficient estimator capacity. Solar geometry controls the smooth daily envelope, which the climatology captures directly. Departures from that envelope arise when clouds, aerosols, ozone, surface albedo, and sensor conditions modulate atmospheric transmission. Because those variables are unavailable, boosting can only infer coarse proxies from humidity, precipitation, and wind. The remaining error should therefore be interpreted as conditional uncertainty under incomplete optical-state observation, not merely as an optimization deficit.

#### *The result is an explanatory benchmark, not an operational forecast*

The best full-day model is valuable because it establishes how much of the signal can be explained without contemporaneous weather forecasts. Its month-hour lookup is transparent, inexpensive, and easy to update, which makes it useful for detecting sensor departures and setting a minimum performance target for future models. It is not, however, a weather forecast in the physical sense: it reproduces the expected calendar envelope rather than predicting clouds. Any operational system should therefore report the baseline as an expected radiation profile and reserve the term forecast for models driven by genuinely future meteorological inputs.

#### *Implications for UV-oriented research*

UV and global solar radiation are related but not equivalent. Published UV reconstructions combine a measured UV target with ozone, radiation, visibility, or satellite products (Junk *et al.*, 2007; Jiang *et al.*, 2024), while multistep UV forecasting studies train directly on UV-index series (Oliveira *et al.*, 2021).

Accordingly, global-radiation predictions may support sensor diagnostics or solar-resource planning, but converting them into UV index requires an independently validated physical or statistical calibration. A future Puno study should collect colocated erythemal irradiance, cloud fraction, ozone, aerosol optical depth, and albedo.

#### *Why meteorological complexity produced limited gains*

The comparison clarifies what the available predictors can and cannot explain. Pressure, temperature, humidity, precipitation, and wind describe the local atmospheric state, but their contemporaneous values are indirect proxies for optical transmission. The month-hour baseline already captures solar geometry and the average seasonal envelope. Boosting can adjust that envelope through nonlinear thresholds, which explains its small daytime RMSE advantage, but it cannot identify every transient cloud field from the supplied columns. Similar environmental studies obtain stronger attribution when they combine meteorology with emissions, satellite retrievals, circulation regimes, or explicit cloud and UV variables (Hsu et al., 2024; Aman et al., 2025; Tong et al., 2025; Wan et al., 2026). The result is therefore not evidence against machine learning; it is evidence that target-relevant measurement matters more than algorithmic complexity.

#### *Reproducibility and operational interpretation*

The temporal split makes the reported errors closer to a deployment question than random cross-validation would: every 2024 prediction is evaluated against observations that occur after the training period. The released package records the raw-file checksum, preprocessing rules, trained climatology, nonlinear model, test predictions, tables, and figures. This design allows another analyst to reproduce the exact results, substitute a different estimator, or append new station years without silently changing the original workbook. Operational use should nevertheless distinguish estimation from forecasting: the meteorological regressors use variables observed at the target hour, so they are diagnostic models unless future weather predictors are supplied by a forecast system.

#### *Limitations and next validation study*

Four limitations bound the conclusions. First, the study uses one station, so spatial generalization across the Lake Titicaca basin is unknown. Second, the file contains no quality flags or maintenance metadata; anomalous pressure and precipitation values were retained rather than removed without documentary justification. Third, positive-radiation hours are defined empirically by the observed target, which is suitable for evaluation but not an operational daylight classifier. Fourth, no direct UV, cloud, ozone, aerosol, or albedo measurements are available. A next study should deploy a calibrated erythemal radiometer alongside the meteorological station, define sensor-quality rules prospectively, evaluate rolling-origin forecasts, and reserve at least one complete year and one external station for validation.

#### *For Puno, the next investment should improve measurement before model complexity*

The evidence points to a concrete priority for local institutions.

A more elaborate algorithm trained on the same columns is unlikely to resolve rapid optical variability, because the missing information is physical rather than computational. The next investment should therefore pair the existing station with calibrated erythemal UV, cloud-fraction, ozone, aerosol, and maintenance records. Once these measurements are available, the current climatology provides a fixed benchmark against which random forests, sequence models, and physically informed hybrids can be evaluated prospectively. This sequence—measure, benchmark, then model—reduces the risk of deploying an impressive but unvalidated public-health indicator.

## CONCLUSION

A prospective evaluation of five years of hourly data shows that Puno's global solar radiation is dominated by stable month-hour structure. The transparent climatology explains 89.6% of 2024 variance and achieves the lowest full-day MAE and RMSE; nonlinear boosting offers only a 1.0% daytime RMSE advantage while increasing typical daytime error. The evidence therefore supports a simple decision: retain the climatology as the benchmark, describe it as an expected radiation profile, and do not infer UV index from it. A defensible public-health warning model requires direct UV and atmospheric optical measurements before additional algorithmic complexity.

**ACKNOWLEDGMENTS:** The authors thank the Vicerrectorado de Investigación of the Universidad Nacional del Altiplano de Puno for institutional support through the Research Seedbeds program, and the Servicio Nacional de Meteorología e Hidrología del Perú for the climatological records analyzed in this study.

**CONFLICT OF INTEREST:** None

**FINANCIAL SUPPORT:** This research was financially supported by the Vice-Rectorate for Research (VRI), Universidad Nacional del Altiplano (UNA Puno), Peru. The funding was provided through the institutional research support program of VRI-UNA Puno. The funder supported the development of the research but had no role in the study design; data collection, analysis, or interpretation; manuscript preparation; or the decision to submit the article for publication.

**ETHICS STATEMENT:** None

## REFERENCES

- Alexander, S. J., & Patel, C. (2025). Latest advancements in snow depth measurement using electromagnetic signals: A review. In S. K. Ghosh, A. Bera, & A. K. Bandyopadhyay (Eds.), *Proceedings of the National Conference on Emerging Trends in Vacuum Electronic Devices and Applications: VEDA2024, 18–20 Nov, Pilani, India* (pp. 342–351). Springer. doi:10.1007/978-981-96-5174-0\_50
- Al-Farsi, M., Al-Harthy, S., & Al-Rawahi, N. (2026). Multimodal AI Copilot for formulation development using protocols, excipient data, and dissolution curves. *Pharmacophore*, 17(1), 43–52. doi:10.51847/ckHNOFgliz

- Aman, R., Rizwan, M., & Kumar, A. (2025). A novel hybrid intelligent approach for solar photovoltaic power prediction considering UV index and cloud cover. *Electrical Engineering*, 107(1), 1203–1224. doi:10.1007/s00202-024-02577-4
- Ashraf, S. F., Jyothi, K., Aejaz, M., & Lohithreddy, A. L. (2025). Weather forecast prediction: An integrated approach for analysing and measuring weather data. In *2025 2nd International Conference on Advanced Computing and Emerging Technologies (ACET 2025)*. IEEE. doi:10.1109/ACET67282.2025.11430270
- Bainapalepu, S. T., & Bathula, K. (2025). Towards safe commute: An AI-driven fusion model for real-time forecasting and emergency alerting. In *2025 IEEE Global Humanitarian Technology Conference (GHTC 2025)*. IEEE. doi:10.1109/GHTC66843.2025.11266147
- Barkani, W., & Allouani, S. A. (2025). Unpacking individual change capacity through a psychological capital: A review of organizational behavior literature. *Journal of Organizational Behavior Research*, 10(3), 28–45. doi:10.51847/SeAEujPy04
- Bedi, P., Das, S., Singh, D. P., & Das, I. (2026). A unified framework for forecast-adaptive and regime-aware ICT design in smart waste collection. *Frontiers in Sustainable Cities*, 8, 1777928. doi:10.3389/frsc.2026.1777928
- Bi, Y., Gao, C., Wang, Y., Pei, J., Zhang, J., & Wang, W. (2026). Correlation between the mesostructural characteristics of asphalt pavement and its gradient aging behavior under intense UV radiation. *Construction and Building Materials*, 509, 145181. doi:10.1016/j.conbuildmat.2026.145181
- Bouali, K., Bajit, A., Benzine, H., Essamri, H., Achour, Y., El Fadi, H., & El Bouayadi, R. (2026). An optimized and intelligent edge computing approach to enhance agricultural sustainability in intelligent farming. In *ICT for Intelligent Systems* (pp. 83–97). Springer. doi:10.1007/978-981-95-1365-9\_8
- Brown, J. N., & Martinez, A. D. (2024). Management of the maize pest *Anomala denuda* (Coleoptera: Scarabaeidae) using an aqueous extract of *Ricinus communis*. *Entomological Research Letters*, 4(2), 1–10. doi:10.51847/62lxylw2KqKI
- Chandrasekhar, K., Nastro, R. A., Narayana, A. V., Rao, A. R., Singh, M., & Chennaiah, A. (2025). Microbial cell factory engineering for scalable production of bio-commodities: Emphasis on robustness. *Journal of Biochemical Technology*, 16(3), 19–29. doi:10.51847/6z0mtOiyyc
- Chen, X., Wang, Z., Su, J., Li, H., Xiong, J., Fu, K., Wang, Z., Yuan, X., Shi, Z., Miao, X., et al. (2023). Altitude-dependent metabolite biomarkers reveal the mechanism of plateau pika adaptation to high altitudes. *Integrative Zoology*, 18(6), 1041–1055. doi:10.1111/1749-4877.12710
- Collins, G., Evans, O., Williams, N., & Brooks, E. (2025). Moral injury, complicity, silence, and withdrawal in organizational life. *Journal of Applied Organizational Systems and Behavior*, 5(2), 61–69. doi:10.51847/xd4sUvLDIS
- Dabeedool, J., Booihawon, R., Gukhool, O., & Shrestha, D. (2023). A comparative study of deep learning algorithms and SARIMA models for forecasting monthly solar radiation and UV index: Case study for Mauritius. In *Lecture Notes in Networks and Systems*, 729 (pp. 273–283). Springer. doi:10.1007/978-3-031-36246-0\_26
- Gong, Q., Xie, L., Dou, D., Wang, K., & Zhang, G. (2022). Prediction of sunburn in power grid workers using machine learning models. In *Proceedings of the 2022 International Conference on Cloud Computing, Big Data Applications and Software Engineering (CBASE 2022)* (pp. 228–231). IEEE. doi:10.1109/CBASE57816.2022.00050
- Hashemzadeh, M., Borhani, F., & Shahverdi, P. (2026). Comparative forecasting of Total Ozone Column over Iran: A 20-year multi-model assessment. *Iran Journal of Computer Science*, 9, 6. doi:10.1007/s42044-025-00388-9
- Hsu, C.-Y., Lee, R.-Q., Wong, P.-Y., Lung, S.-C. C., Chen, Y.-C., Chen, P.-C., Adamkiewicz, G., & Wu, C.-D. (2024). Estimating morning and evening commute period O<sub>3</sub> concentration in Taiwan using a fine spatial-temporal resolution ensemble mixed spatial model with Geo-AI technology. *Journal of Environmental Management*, 351, 119725. doi:10.1016/j.jenvman.2023.119725
- Ianevski, A., Zusinaite, E., Shtaida, N., Kallio-Kokko, H., Valkonen, M., Kantele, A., Telling, K., Tenson, T., & Merits, A. (2019). Low temperature and low UV indexes correlated with peaks of influenza virus activity in northern Europe during 2010–2018. *Viruses*, 11(3), 207. doi:10.3390/v11030207
- Jiang, Y., Shi, S., Li, X., Xu, C., Kan, H., Hu, B., & Meng, X. (2024). A 10 km daily-level ultraviolet-radiation-predicting dataset based on machine learning models in China from 2005 to 2020. *Earth System Science Data*, 16(10), 4655–4672. doi:10.5194/essd-16-4655-2024
- Junk, J., Feister, U., & Helbig, A. (2007). Reconstruction of daily solar UV irradiation from 1893 to 2002 in Potsdam, Germany. *International Journal of Biometeorology*, 51(6), 505–512. doi:10.1007/s00484-007-0089-4
- Kuo, C. Y., Yang, W. W., & Su, E. C. Y. (2024). Improving dengue fever predictions in Taiwan based on feature selection and random forests. *BMC Infectious Diseases*, 24(Suppl. 2). doi:10.1186/s12879-024-09220-4
- Lei, L., Yuan, X., Fu, K., Chen, Y., Lu, Y., Shou, N., Wu, D., Chen, X., Shi, J., Zhang, M., et al. (2022). Pseudotargeted metabolomics revealed the adaptive mechanism of *Draba oreades* Schrenk at high altitude. *Frontiers in Plant Science*, 13, 1052640. doi:10.3389/fpls.2022.1052640
- Li, H., Liu, X., Liu, Z., Li, M., Wu, T., Li, P., & Zhou, P. (2025). Tracking long-term ozone pollution dynamics in Chinese cities with meteorological and emission attribution. *Atmosphere*, 16(7), 768. doi:10.3390/atmos16070768
- Liu, F., Liu, Y., Li, X., Wang, Y., Jing, Z., & Chen, Y. (2025). The analysis of ozone pollution in urban agglomerations of the Sichuan basin based on the scale decomposition-synthesis method. *Environmental Research*, 271, 121121. doi:10.1016/j.envres.2025.121121
- Liu, S., Wang, G., Kong, F., Zhao, N., Gao, W., & Zhang, H. (2025). PM<sub>2.5</sub> pollution characteristics, drivers, and regional transport during different pollution levels in Linyi, China: An integrated PMF-ML-SHAP framework and transport models. *Journal of Hazardous Materials*, 494, 138534. doi:10.1016/j.jhazmat.2025.138534
- Liu, T. C., Shao, X., Choi, T., & Zhang, F. (2025). Long short-term memory (LSTM) deep learning-based prediction of sensor performance time series for improved VIIRS reflective solar band calibration. In *Proceedings of SPIE—The International Society for Optical Engineering*, 13267. SPIE. doi:10.1117/12.3045240

- Liu, X., Song, Y., Yi, W., Wang, X., & Zhu, J. (2018). Comparing the random forest with the generalized additive model to evaluate the impacts of outdoor ambient environmental factors on scaffolding construction productivity. *Journal of Construction Engineering and Management*, 144(6). doi:10.1061/(ASCE)CO.1943-7862.0001495
- Lyu, Y., Wang, Y., Jiang, C., Ding, C., Zhai, M., Xu, K., Wei, L., & Wang, J. (2023). Random forest regression on the joint role of meteorological variables, demographic factors, and policy response measures in COVID-19 daily cases: Global analysis in different climate zones. *Environmental Science and Pollution Research*, 30(32), 79512–79524. doi:10.1007/s11356-023-27320-7
- Marcucci, A., Jerse, G., Alberti, V., & Messerotti, M. (2023). A deep learning model based on multi-head attention for long-term forecasting of solar activity. *Engineering Proceedings*, 39(1), 16. doi:10.3390/engproc2023039016
- Meyer, L., Schmid, A., & Braun, S. (2025). The periodontal envelope of orthodontic tooth movement: Defining anatomical boundaries for safe alignment in reduced alveolar bone conditions. *Annals of Dental Specialty*, 13(2), 120–128. doi:10.51847/NAYqxrjeCQ
- Mohammad, A. A., Chin, O. S., & Elnaem, M. (2024). Understanding diabetes management among patients in Hail City using the health belief model. *Journal of Advanced Pharmacy Education and Research*, 14(4), 28–33. doi:10.51847/AqAVGRxPXc
- Oliveira, P., Fernandes, B., Analide, C., & Novais, P. (2021). Multistep ultraviolet index forecasting using long short-term memory networks. In *Advances in Intelligent Systems and Computing*, 1237 (pp. 187–197). Springer. doi:10.1007/978-3-030-53036-5\_20
- Sanchez, Z., Alva, A., Zimic, M., & Del Carpio, C. (2021). An algorithm for characterizing skin moles using image processing and machine learning. *International Journal of Electrical and Computer Engineering*, 11(4), 3539–3550. doi:10.11591/ijece.v11i4.pp3539-3550
- Scabbia, G., Sanfilippo, A., Mazzoni, A., Bachour, D., Perez-Astudillo, D., Bermudez, V., Wey, E., Marchand-Lasserre, M., & Saboret, L. (2022). Does climate help modeling COVID-19 risk and to what extent? *PLOS ONE*, 17(9), e0273078. doi:10.1371/journal.pone.0273078
- Teklu, A. A., Heckenbach, I., Petr, M. A., Bakula, D., Keijzers, G., & Scheibye-Knudsen, M. (2025). Deep learning reveals diverging effects of altitude on aging. *GeroScience*, 47(3), 3873–3889. doi:10.1007/s11357-024-01502-8
- Tian, N., Zheng, J. X., Li, L. H., Xue, J. B., Xia, S., Lv, S., & Zhou, X. N. (2024). Precision prediction for dengue fever in Singapore: A machine learning approach incorporating meteorological data. *Tropical Medicine and Infectious Disease*, 9(4), 72. doi:10.3390/tropicalmed9040072
- Tong, Z., Yan, Y., Kong, S., Niu, X., & Ma, J. (2025). Improving ozone estimation during rainy-warm seasons from the perspective of weather systems based on machine learning. *Science of the Total Environment*, 958, 177975. doi:10.1016/j.scitotenv.2024.177975
- Wan, Y. Q., Tong, J. L., Liu, Y. L., Wangs, H. S., Yang, H., & Chen, M. (2026). Quantifying the influence of meteorological factors on summer ozone concentration in the main urban area of Lanzhou City based on the XGBoost algorithm. *Environmental Science*, 47(1), 223–232. doi:10.13227/j.hjlx.202412038
- Wang, G., Yang, X., Liu, S., & Kong, F. (2025). Unveiling the drivers of PM2.5 pollution in an industrial inland city in China during heating seasons (2021–2024) by an integrated machine learning method. *Journal of Hazardous Materials*, 500, 140495. doi:10.1016/j.jhazmat.2025.140495
- Wright, T., & Matthews, C. (2025). Prospective analysis of pre-hospital cerebral oxygen saturation monitoring during CPR by emergency personnel. *Journal of Integrative Nursing and Palliative Care*, 6(1), 192–199. doi:10.51847/7rflB3bemK
- Wu, L., Liu, Y., Shi, W., Chang, T., Liu, P., Liu, K., Yong, H., Li, Z., Shi, M., Jiao, N., et al. (2025). Uncovering the hidden RNA virus diversity in Lake Nam Co: Evolutionary insights from an extreme high-altitude environment. *Proceedings of the National Academy of Sciences of the United States of America*, 122(6), e2420162122. doi:10.1073/pnas.2420162122
- Wu, Y., Zhao, G., Li, Y., Li, G., Wang, W., Yang, L., Lin, Z., Yao, H., Jiao, F., Zhao, G., et al. (2026). Metabolomic analysis of Yunnan cigar tobacco leaves: Impact of geography and climate on flavor characteristics and machine learning-based origin traceability. *Frontiers in Plant Science*, 16, 1703429. doi:10.3389/fpls.2025.1703429
- Zhu, X., Hou, J., Li, M., Xu, L., Li, X., Li, Y., Cheng, C., Zhao, W., & He, N. (2023). High-resolution spatial distribution of vegetation biomass and its environmental response on Qinghai-Tibet Plateau: Intensive grid-field survey. *Ecological Indicators*, 149, 110167. doi:10.1016/j.ecolind.2023.110167