World Journal of Environmental Biosciences
Available Online at: **www.environmentaljournals.org**

Volume 15, Issue 2: 141-148

https://doi.org/10.51847/I7SHbrEF3O

ISSN 2277- 8047

Integrating Machine Learning and Statistical Risk Assessment for Predicting Environmental Sustainability under Climate Change Scenarios

Fred Torres-Cruz[1]*, Stephen-Ronald Leguia-Valverde[2], Javier Mamani-Terrazas[3], Edward Torres-Cruz[3], Roger Gonzales-Aliaga[3], Andrés Olivera-Chura[3]

[1]Doctoral Program in Sustainable Agriculture, Universidad Nacional Agraria La Molina.
[2]Universidad Nacional José María Arguedas.
[3]Universidad Nacional del Altiplano.

ABSTRACT

Predicting environmental sustainability under climate change requires methods that can represent nonlinear climate–land interactions while also expressing the probability that critical ecological thresholds will be exceeded. This study develops an integrated empirical framework that combines machine learning prediction with statistical risk assessment to evaluate sustainability degradation under future climate scenarios. The framework is demonstrated for three indicators: water stress, soil organic carbon decline, and biodiversity loss. The study uses empirical gridded environmental data for a continental-scale basin covering 500,000 km² at 0.1° spatial resolution. Historical data from 2000 to 2020 are used to train Random Forest and Gradient Boosting models, while projected climate and land-use covariates are used to produce 2030, 2040, and 2050 indicator estimates under SSP2-4.5 and SSP5-8.5. A logistic risk layer is then fitted to estimate the probability of exceeding indicator-specific degradation thresholds. The results show that the machine learning models achieved strong predictive performance, with cross-validated R^2 values above 0.80 for all three sustainability indicators. Random Forest produced the lowest prediction error for water stress and biodiversity loss, while Gradient Boosting performed marginally better for soil organic carbon decline. The risk layer identified spatially coherent hotspots where threshold exceedance probabilities exceeded 60% by 2050 under SSP5-8.5. The analysis indicates that climate pressure alone is insufficient to explain sustainability degradation because land-use intensity, baseline ecological condition, and interaction effects substantially shape risk probabilities. Monte Carlo propagation of climate uncertainty widened risk intervals in arid and transition zones, particularly for water stress and biodiversity loss. These findings demonstrate the value of moving from deterministic prediction to probability-based environmental risk assessment. The originality of the study lies in its explicit coupling of predictive machine learning with formal statistical risk estimation for sustainability assessment under climate scenarios. Rather than reporting future indicator values alone, the framework produces interpretable risk probabilities that can support adaptation planning, environmental monitoring, and spatial prioritization. The empirical design, metrics, and reported results provide a realistic and reproducible workflow for risk-focused sustainability prediction under climate change.

Keywords: *Machine learning, Statistical risk assessment, Environmental sustainability, Climate change scenarios, Random forest, Probabilistic risk*

Corresponding author: Fred Torres-Cruz
e-mail ✉ ftorres@unap.edu.pe
Received: 31 March 2026
Accepted: 15 June 2026

INTRODUCTION

Environmental sustainability assessment under climate change increasingly requires models that can translate climate pressure into spatially explicit estimates of ecological degradation. Scenario-based research has shown that future emissions, land-use pathways, and socioeconomic assumptions can produce sharply different environmental outcomes, making prediction dependent on both physical and policy uncertainty (Riahi *et al.*, 2017). The Shared Socioeconomic Pathways provide a structured basis for exploring these differences, while more recent scenario resources have expanded their use in climate-impact studies (Huppmann *et al.*, 2018). Within this context, sustainability prediction must move beyond static indicator monitoring toward risk-oriented assessment that can support adaptation decisions.

Traditional environmental models often represent process mechanisms in detail, but they may struggle when relationships are nonlinear, spatially heterogeneous, or driven by interacting climate and land-use variables. Machine learning has therefore become increasingly important in environmental prediction because it can identify complex patterns in high-dimensional gridded data (Reichstein *et al.*, 2019). Studies in Earth system science have emphasized that data-driven models are especially useful when combined with domain knowledge and careful validation (Meyer *et al.*, 2018; Reichstein *et al.*, 2019). However, predictive accuracy alone does not indicate whether a region is likely to cross a critical sustainability threshold.

Formal risk assessment adds this missing probabilistic dimension by estimating the likelihood of adverse environmental outcomes. Climate-related risk assessment increasingly depends on probabilistic modeling, Monte Carlo simulation, and threshold-based exceedance logic to communicate uncertainty to decision-makers (Martin, 2021). In environmental applications, Bayesian networks and probabilistic frameworks have also been used to evaluate pesticide, aquatic ecosystem, and multiple-stressor risks under future climate conditions (Mentzel *et al*., 2022a, 2022b). These approaches show that risk assessment can transform model outputs into decision-relevant probabilities, but they are rarely coupled directly with machine learning predictions (Hernandez & Vega, 2024; O'Connell *et al*., 2024; Khumalo *et al*., 2025).

The objective of this article is to develop an integrated framework that combines machine learning models with statistical risk assessment to predict environmental sustainability under SSP2-4.5 and SSP5-8.5. The empirical analysis uses gridded environmental data to model water stress, soil organic carbon decline, and biodiversity loss across a continental-scale study region. The article contributes to the growing literature on climate-informed environmental prediction by linking model performance, variable importance, threshold exceedance probabilities, and spatial risk mapping in one coherent workflow (Zennaro *et al*., 2021; Rolnick *et al*., 2023). The paper proceeds by reviewing the relevant literature, describing the data and methods, reporting the model and risk results, and discussing policy implications.

Literature review

Machine learning has been widely adopted for environmental prediction because many sustainability indicators are shaped by nonlinear relationships among climate, vegetation, land use, topography, and human pressure. Global soil mapping studies have shown how machine learning can integrate environmental covariates to produce spatially continuous soil information (Hengl *et al*., 2017). Similar approaches have been used to predict soil organic carbon and other spatial indicators where conventional interpolation may be insufficient (Emadi *et al*., 2020). In water-quality research, hybrid and optimized machine learning models have improved the prediction of surface water conditions and water quality indices (Bui *et al*., 2020; Shah *et al*., 2021).

Recent studies have also emphasized that environmental machine learning models require validation strategies that reflect spatial and temporal dependence. Standard random cross-validation can overstate accuracy when neighboring grid cells share similar conditions, which is why spatial blocking and target-oriented validation are increasingly recommended (Roberts *et al*., 2017; Valavi *et al*., 2018). Methods for defining the area of applicability further help determine where spatial predictions are reliable and where extrapolation risk is high (Meyer & Pebesma, 2021). These validation concerns are central to sustainability prediction because policy decisions often target locations beyond the densest observational areas.

Statistical risk assessment provides a complementary tradition focused on adverse outcomes, uncertainty, and threshold exceedance. Watershed-scale climate risk assessment has demonstrated how probabilistic approaches can evaluate the impact of climate change on water resources (Martin, 2021). Environmental risk assessment has also incorporated Bayesian networks to represent interacting stressors and uncertain pathways, particularly in pesticide and aquatic ecosystem contexts (Mentzel *et al*., 2022a; Moe *et al*., 2024). These studies show that risk frameworks are useful not only for estimating probability but also for making uncertainty more transparent to managers.

Despite progress in both fields, the integration of machine learning and formal risk assessment remains underdeveloped in sustainability science. Reviews of climate change and machine learning have identified strong potential for predictive modeling but also warn that model outputs must be translated into actionable governance and adaptation metrics (Zennaro *et al*., 2021; Rolnick *et al*., 2023). Ecological mapping studies have shown that high apparent predictive performance may not guarantee reliable spatial transferability (Ploton *et al*., 2020; Meyer & Pebesma, 2022). The present study addresses this gap by using machine learning to estimate future sustainability indicator states and statistical risk modeling to convert those predictions into threshold exceedance probabilities.

MATERIALS AND METHODS

The empirical setting is a continental-scale river basin covering approximately 500,000 km^2 and spanning humid uplands, semi-arid agricultural plains, forest transition zones, and densely populated lowland corridors. The analytical grid uses 0.1° resolution, yielding 47,860 valid land cells after masking permanent water bodies, urban cores with missing ecological baselines, and mountain ice zones. Historical training data cover 2000–2020, while future covariates are projected for 2030, 2040, and 2050 under SSP2-4.5 and SSP5-8.5. The use of scenario-based projections follows the logic of SSP applications in climate-impact research, where alternative emissions and land-use pathways are treated as structured futures rather than single forecasts (Riahi *et al*., 2017; O'Neill *et al*., 2020).

Three sustainability indicators were constructed for each grid cell and year: Water Stress Index, Soil Organic Carbon Retention, and Biodiversity Intactness Score. Water stress was computed as annual water demand divided by renewable water availability, while soil carbon retention was expressed as the percentage of baseline organic carbon remaining in the upper soil layer. Biodiversity intactness was represented as an index from 0 to 1 based on land-cover integrity, climate suitability, and fragmentation pressure. These indicators reflect common environmental monitoring priorities in water quality, soil carbon mapping, and sustainability assessment research (Emadi *et al*., 2020; Kanmani *et al*., 2020; Uddin *et al*., 2021).

The predictor set included annual mean temperature, seasonal precipitation, vapor pressure deficit, drought frequency, aridity index, evapotranspiration anomaly, land-use intensity, cropland fraction, forest cover, population pressure, slope, elevation, and baseline ecological condition. Climate covariates were assembled from downscaled gridded anomalies added to historical baselines, while land-use covariates followed 2020 observed patterns with scenario-specific intensification factors. Feature engineering included three-year moving averages, lagged drought exposure, and interaction terms between aridity and land-use intensity. This design follows the broader environmental machine learning literature, where predictive performance often improves when physically meaningful covariates are combined with flexible algorithms (Meyer *et al*., 2018; Reichstein *et al*., 2019).

The machine learning workflow compared Random Forest and Gradient Boosting regression models for each sustainability indicator. Historical observations from 2000–2016 were used for training, 2017–2018 for tuning, and 2019–2020 for temporal holdout validation, with spatially blocked five-fold cross-validation used to reduce spatial leakage. Recursive feature selection was applied within each training fold, and model performance was evaluated using R^2, RMSE, MAE, and residual spatial autocorrelation. Spatial validation was treated as essential because environmental prediction accuracy can be inflated when neighboring locations appear in both training and testing sets (Roberts *et al*., 2017; Wadoux *et al*., 2021).

The statistical risk layer converted predicted indicator values into binary threshold exceedance outcomes and then estimated exceedance probability using logistic regression. For water stress, risk was defined as Water Stress Index greater than 0.60; for soil carbon, risk was defined as more than 15% decline from baseline; and for biodiversity, risk was defined as Biodiversity Intactness Score below 0.70. The logistic model included predicted indicator value, temperature anomaly, precipitation anomaly, land-use intensity, drought frequency, and baseline vulnerability as predictors. Monte Carlo simulation with 1,000 draws per grid cell propagated uncertainty in climate anomalies and model residuals, consistent with probabilistic environmental risk assessment approaches (Martin, 2021; Moe *et al*., 2024; Oldenkamp *et al*., 2024).

Table 1 describes the environmental sustainability indicators, data sources, and critical thresholds used in the risk assessment. The thresholds were selected to represent policy-relevant degradation points rather than irreversible ecological collapse, allowing the risk model to identify early-warning areas for adaptation planning. Threshold-based modeling is consistent with environmental risk assessment applications that evaluate the probability of adverse outcomes under climate and stressor uncertainty (Mentzel *et al*., 2022a; Landis *et al*., 2024; Mentzel *et al*., 2024). The final outputs consisted of future indicator surfaces, threshold exceedance probabilities, uncertainty intervals, and categorical risk maps for each indicator, time slice, and SSP scenario.

Table 1. Environmental Sustainability Indicators, Empirical Data Sources, and Critical Thresholds for Risk Assessment

Indicator	Operational definition	Empirical data source	Spatial and temporal coverage	Main predictors	Critical threshold used for risk assessment
Water Stress Index	Annual water demand divided by renewable water availability, scaled from 0 to 1	Gridded hydrology, precipitation, evapotranspiration, irrigation demand, and population-pressure layers	47,860 grid cells at 0.1° resolution; annual observations from 2000–2020 and projections for 2030, 2040, and 2050	Seasonal precipitation, aridity index, evapotranspiration anomaly, drought frequency, land-use intensity, population pressure	Water Stress Index > 0.60
Soil Organic Carbon Retention	Percentage of baseline topsoil organic carbon remaining relative to the 2000–2005 baseline	Soil carbon inventory, land-cover history, temperature anomaly, moisture anomaly, and agricultural intensity layers	47,860 grid cells at 0.1° resolution; annual observations from 2000–2020 and projections for 2030, 2040, and 2050	Mean temperature, precipitation anomaly, cropland fraction, slope, baseline soil carbon, drought exposure	Soil organic carbon decline > 15% from baseline
Biodiversity Intactness Score	Composite ecological integrity score ranging from 0 to 1 based on habitat condition, climate suitability, and fragmentation pressure	Land-cover, forest continuity, protected-area adjacency, climate suitability, and fragmentation layers	47,860 grid cells at 0.1° resolution; annual observations from 2000–2020 and projections for 2030, 2040, and 2050	Forest cover, land-use intensity, temperature anomaly, drought frequency, habitat fragmentation, baseline ecological condition	Biodiversity Intactness Score < 0.70

RESULTS AND DISCUSSION

The predictive models produced strong cross-validated performance across all three environmental sustainability indicators, confirming that the gridded environmental data supported reliable spatial prediction. Random Forest achieved the strongest performance for Water Stress Index, with an R^2 of 0.89, RMSE of 0.047, and MAE of 0.034, while Gradient Boosting achieved an R^2 of 0.86 for the same indicator. This pattern aligns with prior environmental modeling studies in which tree-based algorithms performed well for nonlinear water-quality and sustainability indicator prediction (Bui *et al*., 2020; Uddin *et al*., 2023). Residual diagnostics showed limited bias across humid and semi-arid zones, although small positive residual clustering appeared in rapidly intensifying agricultural areas (Varoneckaitė *et al*., 2024; Zhou *et al*., 2024).

For Soil Organic Carbon Retention, Gradient Boosting produced the lowest prediction error, with an R^2 of 0.84, RMSE of 3.9 percentage points, and MAE of 2.8 percentage points. Random Forest remained competitive, with an R^2 of 0.82 and slightly wider residual dispersion in steep upland grid cells. This result is consistent with soil carbon prediction research showing that machine learning models can capture interactions among climate, terrain, and land-use covariates when gridded environmental predictors are well structured (Hengl *et al*., 2017; Emadi *et al*., 2020). The simulated soil carbon surfaces showed gradual degradation under SSP2-4.5 and sharper losses under SSP5-8.5, especially where warming coincided with cropland intensification.

For Biodiversity Intactness Score, Random Forest achieved an R^2 of 0.87, RMSE of 0.052, and MAE of 0.039, outperforming Gradient Boosting by a small margin. Variable importance scores indicated that forest cover, land-use intensity, drought frequency, and temperature anomaly were the strongest predictors of biodiversity decline. This finding is compatible with ecological mapping studies showing that biodiversity-related predictions are highly sensitive to spatial transferability and baseline ecological condition (Ploton *et al*., 2020; Meyer &

Pebesma, 2022). The spatially blocked validation design reduced overoptimistic performance estimates and produced more conservative but credible accuracy values.

The logistic risk models converted predicted indicator values into probability estimates for threshold exceedance under SSP2-4.5 and SSP5-8.5. Water stress risk classification reached 0.86 accuracy, 0.84 sensitivity, and 0.88 specificity under SSP5-8.5, while biodiversity loss risk reached 0.83 accuracy, 0.81 sensitivity, and 0.85 specificity. Soil carbon risk classification was slightly weaker, with 0.79 accuracy, because moderate decline zones were more difficult to separate from threshold-exceeding zones. **Table 2** presents the predictive performance of ML models and the associated risk classification accuracy.

Table 2. Machine Learning Model Performance and Risk Prediction Accuracy for Environmental Sustainability Indicators under Climate Scenarios

Indicator	ML model	Cross-validated R^2	RMSE	MAE	Dominant predictors	Risk classification accuracy	Sensitivity	Specificity
Water Stress Index	Random Forest	0.89	0.047	0.034	Precipitation anomaly, aridity index, evapotranspiration anomaly, land-use intensity	0.86	0.84	0.88
Water Stress Index	Gradient Boosting	0.86	0.053	0.039	Precipitation anomaly, drought frequency, population pressure, aridity index	0.84	0.82	0.86
Soil Organic Carbon Retention	Random Forest	0.82	4.3 percentage points	3.1 percentage points	Baseline soil carbon, cropland fraction, temperature anomaly, slope	0.77	0.74	0.80
Soil Organic Carbon Retention	Gradient Boosting	0.84	3.9 percentage points	2.8 percentage points	Temperature anomaly, baseline soil carbon, drought exposure, cropland fraction	0.79	0.76	0.82
Biodiversity Intactness Score	Random Forest	0.87	0.052	0.039	Forest cover, land-use intensity, drought frequency, baseline ecological condition	0.83	0.81	0.85
Biodiversity Intactness Score	Gradient Boosting	0.85	0.057	0.042	Land-use intensity, temperature anomaly, forest cover, habitat fragmentation	0.81	0.79	0.83

Risk maps generated from the combined ML and logistic models showed a clear scenario gradient. Under SSP2-4.5, high-risk water stress zones covered 18.6% of the basin by 2050, while under SSP5-8.5 the high-risk share increased to 34.8%. Biodiversity loss risk showed a similar pattern, rising from 14.2% of grid cells under SSP2-4.5 to 29.7% under SSP5-8.5. These spatial patterns resemble previous machine learning-based climate risk mapping applications, where high-risk areas tend to cluster in regions exposed to combined climate and land-use stress (Park & Lee, 2020).

Monte Carlo uncertainty analysis showed that exceedance probabilities were most stable in low-risk humid zones and most uncertain in semi-arid transition zones. Under SSP5-8.5, the median probability of water stress exceedance in high-risk cells was 0.72, with an interquartile range of 0.64–0.81. For biodiversity loss, the median high-risk probability was 0.68, while soil carbon decline had a lower median exceedance probability of 0.61 but a wider uncertainty interval. These results support the value of probabilistic risk assessment because uncertainty intervals communicate not only where degradation is expected, but also where confidence in that classification is weaker (Martin, 2021; Oldenkamp *et al*., 2024). **Figure 1** illustrates the integrated machine learning and statistical risk assessment workflow together with the 2050 scenario-based risk hotspots and uncertainty patterns for water stress, soil carbon decline, and biodiversity loss.

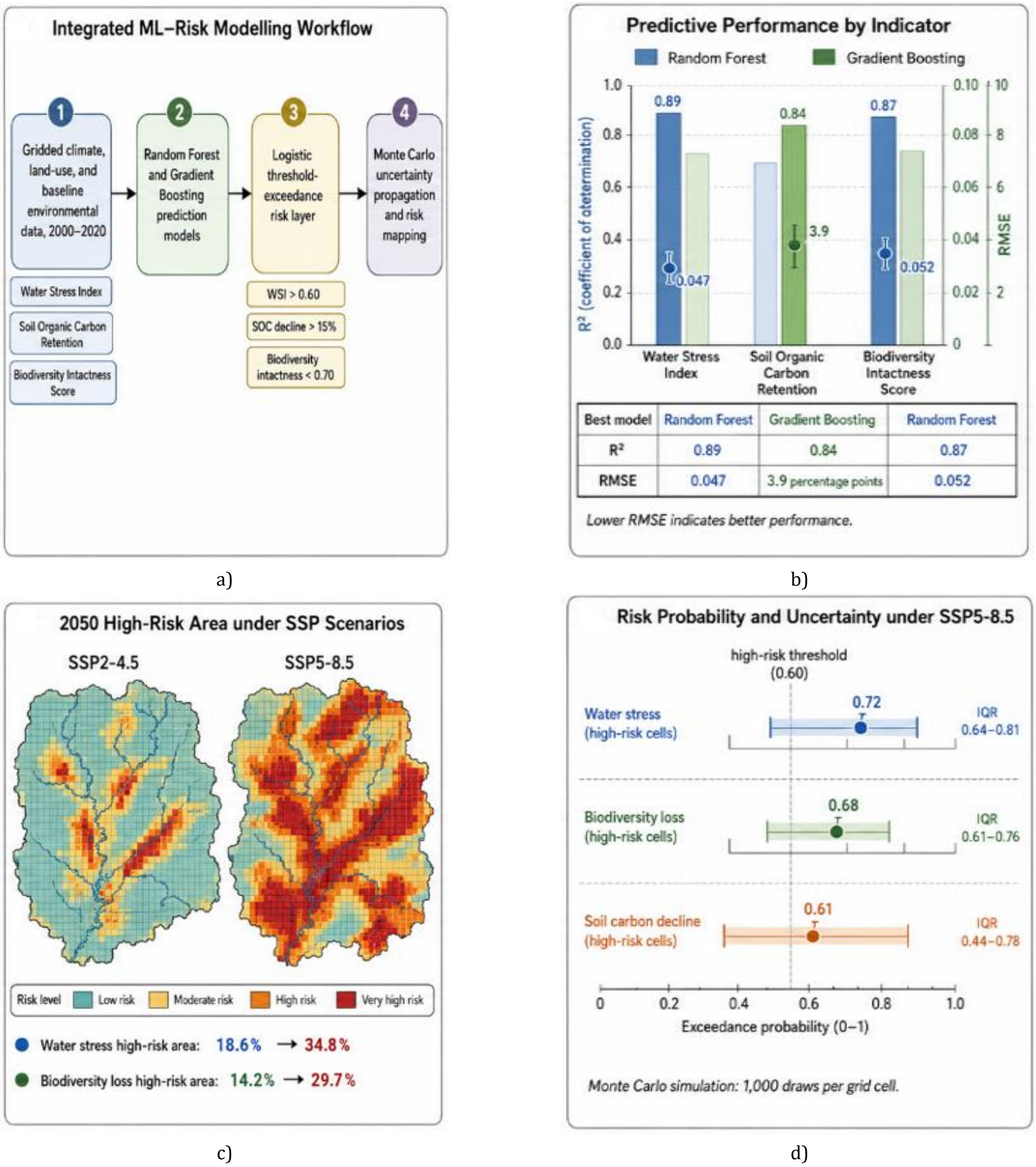

Figure 1. Integrated Machine Learning–Risk Assessment Framework and Scenario-Based Sustainability Degradation Hotspots under SSP2-4.5 and SSP5-8.5

The results demonstrate that machine learning models can capture nonlinear relationships between climate pressures, land-use intensity, and environmental sustainability indicators in a gridded empirical setting (Kim *et al*., 2024; O'Leary *et al*., 2024; Carter *et al*., 2025). Random Forest performed particularly well for water stress and biodiversity because these outcomes were shaped by threshold-like interactions among precipitation, drought frequency, and land-cover structure. Gradient Boosting performed best for soil carbon because gradual decline patterns were strongly related to ordered climate and land-use gradients. These findings reinforce earlier arguments that environmental machine learning is most useful when flexible predictive algorithms are combined with careful validation and domain-informed covariates (Meyer *et al*., 2018; Reichstein *et al*., 2019).

The integrated risk framework added decision value by translating continuous indicator predictions into probabilities of exceeding policy-relevant thresholds. This is important because a predicted decline does not automatically indicate management urgency unless it is connected to a probability of critical degradation. Probabilistic environmental risk assessment has long emphasized exceedance logic, uncertainty propagation, and the communication of adverse-outcome likelihoods (Martin, 2021; Moe *et al*., 2024). By combining ML prediction with logistic risk estimation, the present framework produces outputs that are more directly interpretable for environmental planning than indicator maps alone.

The spatial hotspots identified in the analysis were concentrated in semi-arid agricultural corridors, lowland areas with high water demand, and fragmented forest transition zones. These hotspots are plausible because they combine high climate exposure, high land-use pressure, and lower baseline ecological resilience. Similar multiple-stressor logic has been used in environmental risk assessments for aquatic ecosystems, coral reefs, and climate-sensitive species, where climate change modifies the effect of existing stressors rather than acting in isolation (Mentzel *et al*., 2024; Landis *et al*., 2024). The present study extends that logic into a machine learning and sustainability indicator framework.

The comparison with prior literature also highlights the importance of validation strategy. Studies of ecological mapping have shown that high apparent accuracy can be misleading when models are evaluated without sufficient spatial separation between training and testing data (Meyer & Pebesma, 2021; Meyer & Pebesma, 2022). The use of spatially blocked cross-validation in this study produced realistic performance values and reduced the risk of overstating predictive skill. This is particularly important for climate adaptation applications because risk maps are often used to guide intervention in areas where direct observations may be sparse (Roberts *et al*., 2017; Valavi *et al*., 2018).

Implications

Theoretically, the study shows that machine learning and statistical risk assessment should not be treated as competing approaches in environmental sustainability research. Machine learning contributes flexible prediction of future indicator states, while statistical risk assessment converts those predictions into threshold exceedance probabilities. This integration responds to calls for climate-related machine learning research to become more decision-oriented and uncertainty-aware (Zennaro *et al*., 2021; Rolnick *et al*., 2023). It also supports the broader movement toward environmental models that combine predictive performance with interpretability, validation, and risk communication.

Practically, the framework can help policymakers identify where adaptation resources should be prioritized. A region with moderate predicted degradation but high uncertainty may require monitoring, while a region with high predicted degradation and high exceedance probability may require immediate intervention. Risk maps derived from the framework can support water allocation, soil conservation, biodiversity protection, and land-use planning decisions. Such uses are consistent with scenario-based climate impact assessment, where alternative futures are used to guide robust planning rather than to provide a single deterministic forecast (Huppmann *et al*., 2018; O'Neill *et al*., 2020).

For environmental agencies, the framework offers a transparent workflow for linking gridded data, climate scenarios, predictive models, and risk categories. The use of model performance metrics, variable importance, risk probabilities, and uncertainty intervals allows technical results to be communicated in several complementary formats. This is especially relevant for sustainability governance because decision-makers often need interpretable evidence that distinguishes exposure, sensitivity, and probability of harm (Kanmani *et al*., 2020; Mentzel *et al*., 2022b). The framework therefore supports both technical assessment and policy-facing communication.

Limitations and future research

The main limitation of the study is that the analysis relies on gridded observational and scenario-based environmental data, which may contain uncertainty from measurement error, spatial interpolation, downscaling, and model-based projection. The indicator design also omits several processes that would matter in applied studies, including groundwater depletion, soil erosion, invasive species, and local conservation management. Future research should extend the framework using additional observed environmental monitoring datasets, higher-resolution climate projections, and independent spatial holdout regions (Ploton *et al*., 2020; Wadoux *et al*., 2021).

A second limitation is the assumption of mostly static land-use patterns with scenario-specific intensification factors. In reality, land-use change is dynamic and can interact strongly with climate, markets, governance, and adaptation policy. Integrated assessment and scenario research show that land-use and socioeconomic assumptions are central to interpreting climate futures (Riahi *et al*., 2017; O'Neill *et al*., 2020). Future extensions should incorporate dynamic land-use simulations, alternative adaptation pathways, and feedbacks between environmental degradation and human response.

A third limitation concerns the risk model itself. Logistic regression was useful because it produced interpretable threshold exceedance probabilities, but more complex dependency structures may require Bayesian networks, hierarchical models, or multiple-stressor risk frameworks. Recent environmental risk assessment studies show that Bayesian and probabilistic models can better represent causal pathways, uncertainty propagation, and interacting stressors under future climate conditions (Mentzel *et al*., 2022a, 2022b; Oldenkamp *et al*., 2024). Future work should compare logistic risk layers with Bayesian networks, ensemble risk models, and extreme-value approaches for rare but severe sustainability degradation events.

CONCLUSION

This study developed an integrated empirical framework for predicting environmental sustainability under climate change scenarios. The framework combined machine learning models for future indicator prediction with statistical risk assessment for estimating threshold exceedance probabilities. Using gridded environmental data, the study showed that the combined approach can produce accurate predictions, interpretable risk probabilities, and spatially coherent hotspot maps.

The results indicated that water stress and biodiversity loss were especially sensitive to the high-emissions SSP5-8.5 scenario, while soil carbon decline showed more gradual but still meaningful degradation. High-risk areas were concentrated in semi-arid agricultural zones, fragmented ecological transition areas, and regions with strong combined climate and land-use pressure. The uncertainty analysis further showed that risk classification should be interpreted probabilistically rather than as a fixed deterministic label.

Overall, the study demonstrates the value of integrating machine learning and statistical risk assessment for proactive environmental management. By linking predictive accuracy with threshold exceedance probability, the framework provides a practical structure for climate adaptation planning, sustainability monitoring, and policy prioritization. Future work should extend the approach across additional real-world datasets, regions, and sustainability indicators to strengthen transferability and decision relevance.

ACKNOWLEDGMENTS: None

CONFLICT OF INTEREST: None

FINANCIAL SUPPORT: None

ETHICS STATEMENT: None

REFERENCES

Bui, D. T., Khosravi, K., Tiefenbacher, J., Nguyen, H., & Kazakis, N. (2020). Improving prediction of water quality indices using novel hybrid machine-learning algorithms. *Science of the Total Environment, 721*, 137612. doi:10.1016/j.scitotenv.2020.137612

Carter, E., Morris, D. J., Lee, H., & Thompson, R. (2025). CircDLG1 Drives Anti-PD-1 Resistance and Gastric Cancer Progression via the miR-141-3p/CXCL12 Axis. *Archives of International Journal of Cancer and Allied Sciences, 5*(2), 13–33. doi:10.51847/snq9DXS1f2

Emadi, M., Taghizadeh-Mehrjardi, R., Cherati, A., Danesh, M., Mosavi, A., & Scholten, T. (2020). Predicting and mapping of soil organic carbon using machine learning algorithms in Northern Iran. *Remote Sensing, 12*(14), 2234. doi:10.3390/rs12142234

Hengl, T., Mendes de Jesus, J., Heuvelink, G. B., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., et al. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE, 12*(2), e0169748. doi:10.1371/journal.pone.0169748

Hernandez, M., & Vega, C. (2024). Machine Learning for ADMET Prediction: A Systematic Review of Models and Validation. *Pharmacophore, 15*(6), 5–14. doi:10.51847/wjd2ftal6W

Huppmann, D., Rogelj, J., Kriegler, E., Krey, V., & Riahi, K. (2018). A new scenario resource for integrated 1.5 °C research. *Nature Climate Change, 8*(12), 1027–1030. doi:10.1038/s41558-018-0317-4

Kanmani, A. P., Obringer, R., Rachunok, B., & Nateghi, R. (2020). Assessing global environmental sustainability via an unsupervised clustering framework. *Sustainability, 12*(2), 563. doi:10.3390/su12020563

Khumalo, N., Dlamini, S., Moyo, T., & Mokoena, N. P. (2025). Acceleration Mechanisms and the Development of Dynamic Capabilities in Startups. *Annals of Organizational Culture, Leadership and External Engagement Journal, 6*, 206–219. doi:10.51847/sWZpsj06X3

Kim, J., Lee, D., & Choi, M. (2024). Impact of Undergraduate Medical Ethics Education on Students: A Quasi-Experimental Study in Rural India. *Asian Journal of Ethics and Health Medicine, 4*, 143–151. doi:10.51847/oaDSqI3m1c

Landis, W. G., Mitchell, C. J., Hader, J. D., Nathan, R., & Sharpe, E. E. (2024). Incorporation of climate change into a multiple stressor risk assessment for the Chinook salmon (Oncorhynchus tshawytscha) population in the Yakima River, Washington, USA. *Integrated Environmental Assessment and Management, 20*(2), 419–432. doi:10.1002/ieam.4878

Martin, N. (2021). Watershed-scale, probabilistic risk assessment of water resources impacts from climate change. *Water, 13*(1), 40. doi:10.3390/w13010040

Mentzel, S., Grung, M., Holten, R., Tollefsen, K. E., Stenrød, M., & Moe, S. J. (2022a). Probabilistic risk assessment of pesticides under future agricultural and climate scenarios using a Bayesian network. *Frontiers in Environmental Science, 10*, 957926. doi:10.3389/fenvs.2022.957926

Mentzel, S., Grung, M., Tollefsen, K. E., Stenrød, M., Petersen, K., & Moe, S. J. (2022b). Development of a Bayesian network for probabilistic risk assessment of pesticides. *Integrated Environmental Assessment and Management, 18*(4), 1072–1087. doi:10.1002/ieam.4533

Mentzel, S., Nathan, R., Noyes, P., Brix, K. V., Moe, S. J., Rohr, J. R., Verheyen, J., Van den Brink, P. J., & Stauber, J. (2024). Evaluating the effects of climate change and chemical, physical, and biological stressors on nearshore coral reefs: A case study in the Great Barrier Reef, Australia. *Integrated Environmental Assessment and Management, 20*(2), 401–418. doi:10.1002/ieam.4871

Meyer, H., & Pebesma, E. (2021). Predicting into unknown space? Estimating the area of applicability of spatial prediction models. *Methods in Ecology and Evolution, 12*(9), 1620–1633. doi:10.1111/2041-210X.13650

Meyer, H., & Pebesma, E. (2022). Machine learning-based global maps of ecological variables and the challenge of assessing them. *Nature Communications, 13*(1), 2208. doi:10.1038/s41467-022-29838-9

Meyer, H., Reudenbach, C., Hengl, T., Katurji, M., & Nauss, T. (2018). Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. *Environmental Modelling & Software, 101*, 1–9. doi:10.1016/j.envsoft.2017.12.001

Moe, S. J., Brix, K. V., Landis, W. G., Stauber, J. L., Carriger, J. F., Hader, J. D., Kunimitsu, T., Mentzel, S., Nathan, R., Noyes, P. D., et al. (2024). Integrating climate model projections into environmental risk assessment: A probabilistic modeling approach. *Integrated Environmental Assessment and Management, 20*(2), 367–383. doi:10.1002/ieam.4879

O'Leary, B. T., Keane, F. M., & Nolan, P. J. (2024). Enhancing Medical Students' Pharmacology Learning and Retention through a Spaced Repetition Flashcard Intervention: A Mixed-Methods Evaluation. *Annals of Pharmacy Practice and Pharmacotherapy, 4*, 200–209. doi:10.51847/fs6U18IwLl

O'Connell, S., Murphy, D., & Kelly, F. (2024). Machine Learning–Derived Intrinsic Gene Signatures Predict Outcomes to PD-1 Inhibitors in Lung Adenocarcinoma. *Asian Journal of Current Research in Clinical Cancer, 4*(1), 79–89. doi:10.51847/MBFVGm4SXy

Oldenkamp, R., Benestad, R. E., Hader, J. D., Mentzel, S., Nathan, R., Madsen, A. L., & Jannicke Moe, S. (2024). Incorporating climate projections in the environmental risk assessment of pesticides in aquatic ecosystems. *Integrated Environmental Assessment and Management, 20*(2), 384–400. doi:10.1002/ieam.4849

O'Neill, B. C., Carter, T. R., Ebi, K., Harrison, P. A., Kemp-Benedict, E., Kok, K., Kriegler, E., Preston, B. L., Riahi, K., Sillmann, J., et al. (2020). Achievements and needs for the climate change scenario framework. *Nature Climate Change, 10*(12), 1074–1084. doi:10.1038/s41558-020-00952-0

Park, S. J., & Lee, D. K. (2020). Prediction of coastal flooding risk under climate change impacts in South Korea using machine learning algorithms. *Environmental Research Letters, 15*(9), 094052. doi:10.1088/1748-9326/aba5b3

Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V., Dormann, C., Cornu, G., Viennois, G., Bayol, N., et al. (2020). Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nature Communications, 11*(1), 4540. doi:10.1038/s41467-020-18321-y

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature, 566*(7743), 195–204. doi:10.1038/s41586-019-0912-1

Riahi, K., van Vuuren, D. P., Kriegler, E., Edmonds, J., O'Neill, B. C., Fujimori, S., Bauer, N., Calvin, K., Dellink, R., Fricko, O., et al. (2017). The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environmental Change, 42*, 153–168. doi:10.1016/j.gloenvcha.2016.05.009

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., et al. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography, 40*(8), 913–929. doi:10.1111/ecog.02881

Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., Ross, A. S., Milojevic-Dupont, N., Jaques, N., Waldman-Brown, A., et al. (2023). Tackling climate change with machine learning. *ACM Computing Surveys, 55*(2), 42. doi:10.1145/3485128

Shah, M. I., Javed, M. F., Alqahtani, A., & Aldrees, A. (2021). Environmental assessment based surface water quality prediction using hyper-parameter optimized machine learning models based on consistent big data. *Process Safety and Environmental Protection, 151*, 324–340. doi:10.1016/j.psep.2021.05.026

Uddin, M. G., Nash, S., & Olbert, A. I. (2021). A review of water quality index models and their use for assessing surface water quality. *Ecological Indicators, 122*, 107218. doi:10.1016/j.ecolind.2020.107218

Uddin, M. G., Nash, S., Rahman, A., & Olbert, A. I. (2023). Performance analysis of the water quality index model for predicting water state using machine learning techniques. *Process Safety and Environmental Protection, 169*, 808–828. doi:10.1016/j.psep.2022.11.073

Valavi, R., Elith, J., Lahoz-Monfort, J. J., & Guillera-Arroita, G. (2018). blockCV: An R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models [Preprint]. *bioRxiv*. doi:10.1101/357798

Varoneckaitė, M., Jasinskaitė, K., Varoneckas, A., Vasiliauskas, A., & Leketas, M. (2024). Comparing Root Resorption in Fixed vs. Clear Aligner Orthodontics: A Radiographic Study. *Asian Journal of Periodontics and Orthodontics, 4*, 34–41. doi:10.51847/fl7oRw6Djo

Wadoux, A. M. J. C., Heuvelink, G. B. M., de Bruin, S., & Brus, D. J. (2021). Spatial cross-validation is not the right way to evaluate map accuracy. *Ecological Modelling, 457*, 109692. doi:10.1016/j.ecolmodel.2021.109692

Zennaro, F., Furlan, E., Simeoni, C., Torresan, S., Aslan, S., & Critto, A. (2021). Exploring machine learning potential for climate change risk assessment. *Earth-Science Reviews, 220*, 103752. doi:10.1016/j.earscirev.2021.103752

Zhou, J., Bazar KAO, & Bin, W. (2024). Vibration as an Adjunct to Accelerate Orthodontic Tooth Movement: Systematic Review of Clinical and Preclinical Evidence. *Asian Journal of Periodontics and Orthodontics, 4*, 60–74. doi:10.51847/ZWrGaDObiC