A Survey of AI and IoT-Based Approaches for Climate-Smart Rice Disease and Yield Prediction

Dr. Santhosh Kumar S[1], Dr. Abeer Aljohani[2]

[1]PostDoctoral Researcher, Lincoln University College Malaysia; [2]Department of Computer Science and Information, Applied College, Taibah University, Madinah, Saudi Arabia

pdf.santhoshkumars@lincoln.edu.my; aahjohani@taibahu.edu.sa

Abstract: Climate variability poses an increasing risk for rice cultivation while operational climate products are still too coarse (tens to hundreds of kilometres) to be used for making decisions on field level regarding diseases and yields. Despite significant progress in AI-IoT disease detection, the disease detection systems are primarily image-based and are independent from the climatic drivers, and digital twin (DT) applications have been limited to the monitoring and synchronising of data in agriculture so far, but not closed-loop forecasting. This paper introduces an AI-IoT enabled spatial modelling framework of rice disease dynamics and yield prediction based on machine-learning climate downscaling, using a unified architecture that synchronises them. We review the latest literature on both agricultural digital twins and ML-based climate downscaling, as well as AI-IoT rice disease detection, to find the gap between them that motivates this work, and introduce the proposed five-layer framework, goals, and scope. The framework is designed to be a decision-support architecture applicable to climate-smart rice farming, especially in data and resource-limited rice production areas.

Keywords: Digital Twin; Climate Downscaling; Machine Learning; Rice Disease Dynamics; AI-IoT; Precision Agriculture; Yield Prediction; Climate-Smart Agriculture.

Introduction

Rice serves as a staple food for over half of the world's population, and the production of this crop is now becoming increasingly vulnerable to climate variability due to changes in rainfall patterns, extreme temperatures and expansion of favorable conditions for fungal/bacterial pathogens [1, 2]. Rice impacts assessments have always revealed that the relevant variables affecting rice yield – the onset of monsoon rainfall, the temperature during the flowering stage and the disease pressure related to humidity – have spatial scales much smaller than the spatial scale of the coarse outputs of Global Climate Models (GCMs) and reanalysis products usually available to agricultural planners [6].

To overcome this resolution gap, statistical and machine-learning-based downscaling methods have been proposed, such as autoencoders, transformers, and generative adversarial networks (GANs) [3, 4]. These techniques have been demonstrated to perform well when compared to linear baselines in rice yield forecasting applications in low-resource environments, but are often developed and evaluated as individual climate products, not as part of a disease or yield forecasting system.

Concurrently, there has been a rapid development of AI-IoT rice disease detection systems, whose convolutional neural networks and IoT-based sensor/camera networks have been developed, classified, and, in some cases, spatially localized and quantified within near real time [5, 7]. Yet, the conventional approach is one of static image classification at a particular time, while systems that attempt to model the spatial evolution of disease risk over time in response to fine-resolution climate drivers are relatively scarce.

It has been suggested that digital twin may be the architecture to connect these aspects, by being an up-to-date virtual environment representation of the field, combining climate, sensor and imagery information for simulation and decision making [1], [2]. However, as recent reviews have suggested, existing agricultural DT applications are still in the data collection and synchronisation phase and not in closed-loop forecasting [2].

This paper tackles this gap in integration. We propose a DT-driven framework where a climate downscaling module based on machine learning provides climate drivers at field-level of resolution to an AI-IoT based spatial disease-dynamics module, which in turn feeds into a yield prediction module, all in one DT.

Related Work

A. Digital Twins in Agriculture

Current reviews suggest that DT could be considered as an integrating layer above the sensor and satellite and weather data in crop monitoring, pest/disease early warning, and precision operations [1] and [2]. The work done by Awais et al. [1] reviews DT and smart-farming technologies, with a focus on their ability to simulate different scenarios, including variety switching and changed planting dates for adaptation to climate change. In a comprehensive technology review, Zhang et al. [2] point out that the current focus is not on sensing but on model fidelity – the DT is only as useful as the models that are able to work with the data it contains – and that most deployed systems are not forecasting all parts of the production lifecycle, just data collection and synchronisation. Neither review reports an internally performed climate downscaling DT architecture.

B. Machine-Learning-Based Climate Downscaling

Kow et al. [3] present an autoencoder-transformer model specifically designed for smart agriculture weather down-scaling applications, which shows that deep generative models are able to reconstruct field-specific climate surfaces from coarse input. Today, a conditional GAN is used to downscale ERA5 reanalysis data as shown by Manco et al. [4], which marks the current development towards a generative approach to downscaling instead of a strictly regression-based one. The rice-yield context is particularly of interest, as a statistical-downscaling-plus-ML-ensemble study in Punjab indicates that a 15–20% higher $R^2 = 0.791$ and lower MAE = 0.215 is achieved with an ML-ensemble compared to linear baselines, based on CMIP climate projections [6]; and a study in northern Italy that correlates fine-resolution optical data from Sentinel-2 with nitrogen-management decisions is reported for early-season rice yield prediction [5]. An important counterpoint is provided by Fofanah [8] who found, in the context of data limited in Sierra Leone, that models trained on crop statistics alone did not outperform naive persistence, and only the inclusion of satellite-derived climate covariates (CHIRPS rainfall, NASA POWER temperature) resulted in a material error reduction (RMSE 284 vs. 428 kg/ha), meaning that skill is fundamentally limited by the availability of downscaled/observed climate data, rather than by model architecture.

C. AI-IoT Rice Disease Detection and Spatial Modelling

Many studies have used CNNs and other deep learning architectures for rice leaf classification, with high accuracy reported for diseases like blast, bacterial blight, brown spot and tungro [7]. This can be furthered by the integration of IoT into the framework, with the aim of quantifying the severity of the disease, rather than simply a binary yes/no indicator, by estimating the amount of leaf tissue that is infected [7]. Spatial mapping of infected plants in real-time in a field has also been realized using UAV and GPS with IoT systems [9]. What is less explored to date is the integration of these spatial disease-detection pipelines and downscaled climate covariates as explicit drivers of disease spread, as opposed to each field observation being an independent classification task.

Table 1 Summary of Landscape Based on the Four Core Capabilities of the Proposed Framework: Identifying the Integration Gap.

Study / System	Digital Twin	Climate Downscaling	Disease Spatial Modeling	Yield Prediction
Awais et al. [1] – DT/smart farming review	Yes	No	Partial	No
Zhang et al. [2] – DT in agriculture review	Yes	No	No	No
Kow et al. [3] – AI weather downscaling	No	Yes	No	No

Manco et al. [4] – GAN downscaling (ERA5)	No	Yes	No	No
Sentinel-2 up/downscaling rice study [5]	No	Partial	No	Yes
CMIP-ML ensemble rice yield study [6]	No	Yes	No	Yes
IoT-CNN rice/potato disease system [7]	No	No	Partial	No
Fofanah [8] – ML rice yield, data-scarce	No	Partial	No	Yes
This work (proposed)	Yes	Yes	Yes	Yes

Table 1. Comparison of representative related work against digital twin integration, climate downscaling, disease spatial modelling, and yield prediction capabilities.

Proposed Framework (Design)

The proposed framework is structured as a 5-layered digital twin framework:

Layer 1: Physical/Data Layer – in-field IoT sensors (soil moisture, canopy microclimate), UAV/satellite imagery, ground weather stations and coarse resolution GCM/reanalysis climate inputs.

Layer 2 – Climate Downscaling Module: a ML model such as transformer or GAN based model (as per [3] and [4]) to transform coarse climate information to field-resolution climate surfaces (temperature, humidity, rainfall) for the target rice-growing region.

Layer 3 – Disease Dynamics and Spatial Modelling Module: AI-IoT fusion model (CNN/LSTM or spatiotemporal graph-based) for combining downscaled climate surfaces, sensor data, and images to estimate the probability of onset and spread of disease in a spatial grid of the field.

Layer 4 – Yield Prediction Module: combines the disease-risk surface, agronomic and climate parameters to provide a spatially explicit yield estimate.

Layer 5 – Digital Twin Synchronisation and Decision-Support Layer: keeps a continuously up to date virtual reality model of the field and presents alerts/recommendations to the end user.

Objectives

1. Build climate downscaling module for field-resolution climate surfaces based on machine learning from coarse scale data from GCM/Reanalysis for rice-growing regions.

2. Design an AI-IoT spatial modelling module to marry the downscaled climate drivers with sensor and imagery data for predicting onset and spread of rice disease.

4. Embed the disease-risk output in a yield prediction module and integrate both in a digital twin framework.

5.Develop a validation strategy with the available literature (e.g., [6] and [8]) at a case-study rice-growing area.

Scope

The framework is focused on rice as the focal crop and was designed for situations where we cannot find dense in-situ climate networks, but coarse-resolution climate data can be found, a situation that is more commonly true in the many regions where rice cultivation is dominant. To build on this, the present paper focuses on framework design and literature positioning, the empirical validation (dataset selection, model training, and an accuracy assessment in the field) is planned as a next step in this research and will not be reported in the present paper. The post-harvest and supply chain aspects are not included in this scenario; only the major rice pathogens that are climate sensitive are included.

Key Contribution

- A framework that treats climate as an external, coarse-resolution input instead of directly embedding an ML-based climate downscaling layer as a first-class, modelled component, in contrast to previous agricultural DT studies and implementations [1], [2].

- The gap identified in [5], [7] is filled by directly linking downscaled climate output into a spatial (rather than just image-classification) rice disease-dynamics model.

- An end-to-end design that, to the best of our knowledge, is not documented as an integrated system in the reviewed literature (Table 1) and links downscaling → disease dynamics → yield prediction within a single digital twin synchronization layer.

- A design that explicitly takes into account and responds to the data-scarcity caution brought up by [8], which states that forecast skill advances in rice-yield prediction are often driven by the quality of climate covariates rather than model architecture alone.

Discussions

Two complementing lessons for the suggested framework are suggested by the evaluated literature. First, downscaling quality is likely to be the binding constraint on downstream disease and yield forecasting skill. This is consistent with the finding in [8] that the material forecast-error reduction in a data-constrained setting was produced by climate covariates rather than statistical model choice, and with the significant accuracy gains reported when GAN/transformer-based downscaling is used over linear baselines [3], [4], [6]. This suggests that in order to assign failure modes to the appropriate module, evaluation of the suggested framework should provide downscaling accuracy as a separate, upstream parameter rather than just end-to-end yield or disease accuracy.

Second, a natural interface for coupling with downscaled climate surfaces is provided by the disease-detection literature's shift from static classification to severity quantification and spatial localization [7]. However, this coupling has not yet been shown at the system level in the sources examined here. Propagating downscaling uncertainty thru the disease and yield modules without inflating error is a non-trivial design problem that should be explicitly addressed (e.g., via uncertainty quantification passed between layers) in the next phase of this work. This is both the fundamental novelty claim of the framework and its main technical risk.

The current paper's limitations include the fact that it is a design-stage contribution; no empirical accuracy figures for the suggested framework are provided here; and only after Objective 4 (case-study validation) is completed can the framework's performance in comparison to the benchmarks cited (e.g., $R^2 = 0.791$ in [6]; RMSE reduction in [8]) be determined. Although it covers contemporary (2024–2026) sources, the literature search that underpins Table 1 and the related-work synthesis should be regarded as indicative rather than exhaustive, and authors are urged to cross-reference all cited works with their original publications prior to submission.

Conclusions

1. Problem addressed/motivation: The limited coupling of AI-IoT disease-detection systems to climate causes and the low resolution of available climate data hinder field-level rice disease and yield predictions.

2. Method: A five-layer digital twin structure is suggested that combines yield prediction, AI-IoT spatial disease-dynamics modeling, and ML-based climate downscaling into a synchronized architecture.

3. The proposed architecture is motivated by a key result from the literature synthesis: no system under study unifies digital twin orchestration, climate downscaling, disease spatial modeling, and yield prediction simultaneously (Table 1).

4. Limitations and future work: only design and literature positioning are presented in this publication; the immediate next step is empirical validation on a case-study rice-growing region with specific reporting of downscaling accuracy as a separate parameter.

References

[1] M. Awais, X. Wang, S. Hussain, F. Aziz, and M. Q. Mahmood, "Advancing precision agriculture through digital twins and smart farming technologies: A review," AgriEngineering, vol. 7, no. 5, 2025.

[2] R. Zhang, H. Zhu, Q. Chang, and Q. Mao, "A comprehensive review of digital twins technology in agriculture," Agriculture, vol. 15, no. 9, 2025.

[3] P.-Y. Kow et al., "AI-driven weather downscaling for smart agriculture using autoencoders and transformers," Computers and Electronics in Agriculture, vol. 232, art. 110129, 2025.

[4] I. Manco, W. Riviera, A. Zanetti, M. Briscolini, P. Mercogliano, and A. Navarra, "A new conditional generative adversarial neural network approach for statistical downscaling of the ERA5 reanalysis over the Italian Peninsula," Environmental Modelling & Software, vol. 188, art. 106427, 2025.

[5] "Upscaling and downscaling approaches for early season rice yield prediction using Sentinel-2 and machine learning for precision nitrogen fertilisation," Computers and Electronics in Agriculture (ScienceDirect), 2024.

[6] "Predicting rice yield and impact of climate change on rice production using machine learning models," preprint/manuscript, ResearchGate, 2025.

[7] "IoT integrated CNN framework for automated detection and quantification of rice and potato crop diseases," Scientific Reports, 2025.

[8] I. D. Fofanah, "Can Machine Learning Forecast Rice Yields in Data-Constrained Settings? Satellite Climate Data, National Crop Statistics, and Lessons from Sierra Leone," arXiv:2606.13959, 2026.

[9] "A comprehensive review of deep learning approaches for rice disease detection: Datasets, methodologies, and future directions," ScienceDirect, 2025.

[10] "Machine Learning Frameworks for Downscaling Climate and Environmental Variables: A Comprehensive Review," Springer, 2025.