Remote Sensing Applications: Society and Environment 38 (2025) 101613

Contents lists available at ScienceDirect

Remote Sensing Applications: Society and Environment

journal homepage: www.elsevier.com/locate/rsase

Multi-modal data fusion and deep ensemble learning for accurate crop yield prediction

Akshay Dagadu Yewle, Laman Mirzayeva, Oktay Karakuş *

School of Computer Science and Informatics, Cardiff University, Abacws, Senghennydd Road, Cardiff, CF24 4AG, UK

ARTICLE INFO

Keywords:
Precision agriculture
Data fusion
Remote sensing
Machine learning algorithms
Deep learning models
Ensemble methods

ABSTRACT

This study introduces *RicEns-Net*, a novel deep ensemble model for rice yield prediction in the Mekong Delta region of Vietnam, integrating diverse data sources through multi-modal data fusion. The model leverages synthetic aperture radar (SAR), optical remote sensing (Sentinel-1/2/3) and meteorological measurements (surface temperature, rainfall) to improve prediction precision. A comprehensive feature selection reduced over 100 potential predictors to 15 key features across 5 data modalities, mitigating the "*curse of dimensionality*" where the initial field data were acquired through Ernst & Young's (EY) Open Science Challenge 2023. RicEns-Net outperforms previous state-of-the-art models (including winners of the EY Open Science Challenge), achieving a mean absolute error (MAE) of 336 kg/Ha, roughly 5%–6% of the lowest regional yield, and a high R^2, indicating robust predictive capability. These results underscore the benefit of deep ensembles in precision agriculture and demonstrate the potential of multi-modal data integration for more accurate crop yield forecasting.

1. Introduction

This paper is grounded in the purpose and drive behind one of the Sustainable Development Goals (SDG) outlined by the United Nations; a comprehensive set of 17 objectives to be achieved by 2030 (United Nations, 2023b). These goals collectively embody humanity's pursuit of a sustainable future for both the planet and its inhabitants. Serving as a global framework, the 17 SDGs guide international endeavours to address the challenges posed by climate change while harmonising human ambitions for prosperity and improved quality of life. Central to this vision is ensuring food security (United Nations, 2023a). This is particularly important for a significant portion of the global population living in environmentally vulnerable areas affected by climate and weather fluctuations.

Rice is one of the most important staple foods globally, feeding more than half of the world's population. It is cultivated on over 160 million hectares, producing around 500 million metric tons annually, with the majority of production concentrated in Asia. The region accounts for nearly 90% of global rice output, with countries such as China, India, and Indonesia leading in production. Besides being a dietary staple, rice also supports the livelihoods of millions of farmers, playing a vital role in rural economies. However, rice cultivation is highly resource-intensive, requiring significant amounts of water and labour, and is particularly vulnerable to climate change. Rising temperatures, changing precipitation patterns, and the increasing frequency of extreme weather events threaten rice yields globally, posing challenges to food security in many regions.

This study underscores the critical importance of rice crops and Vietnam, focusing on their global and regional significance. Rice, often referred to as a "Gift of God" for its nutritional value, is a cornerstone of food security and public health worldwide. Despite its critical role, less than 8% of global rice production enters international trade (Food and Agriculture Organization (FAO),

* Corresponding author.
E-mail address: karakuso@cardiff.ac.uk (O. Karakuş).

https://doi.org/10.1016/j.rsase.2025.101613
Received 9 February 2025; Received in revised form 5 May 2025; Accepted 28 May 2025
Available online 14 June 2025

2022). This highlights its largely localised consumption patterns, as reported by the Food and Agriculture Organisation Corporate Statistical Database (FAOSTAT). However, Vietnam stands as an exception, being one of the world's leading rice exporters. The Mekong Delta, in particular, is central to the nation's status as a global rice powerhouse, with its fertile paddy fields supporting diverse rice varieties.

In Vietnam, rice holds unparallelled cultural, economic, and ecological importance. It serves as a dietary staple, forming the foundation of traditional cuisine, while also being integral to the livelihoods of millions of rural farmers. The sector is a driver of economic development, supported by government initiatives aimed at boosting productivity and sustainability. However, the ecological and climatic demands of rice cultivation present significant challenges. As a water-intensive crop, rice is highly sensitive to temperature fluctuations, requiring an optimal range of 25–35 °C. Brief exposure to extreme heat or untimely rainfall during the reproductive stage can result in sterility or yield loss. This vulnerability, coupled with the long growing periods, makes rice cultivation particularly susceptible to climate change, underscoring the urgency of sustainable practices in Vietnam and globally.

This paper introduces *RicEns-Net*, a novel ensemble learning model designed for accurate crop yield prediction. By integrating five diverse remote sensing data sources, Sentinel-1, Sentinel-2, Sentinel-3, NASA's Goddard Earth Sciences (GES) Data and Information Services Centre (DISC) and field measurements, the proposed framework leverages sophisticated data engineering techniques to improve predictive performance. The model effectively combines synthetic aperture radar (SAR), multispectral imaging (MSI), meteorological parameters, and ground-truth observations, ensuring a robust and comprehensive approach to crop yield estimation. The inclusion of multi-modal data sources helps mitigate uncertainties associated with individual datasets, improving both spatial and temporal resolution in predictions.

Main contributions

This study makes the following key contributions:

1. We develop a comprehensive multi-modal data fusion strategy by integrating synthetic aperture radar (SAR), optical remote sensing, and meteorological measurements, significantly enhancing the accuracy of crop yield prediction.
2. We propose *RicEns-Net*, a novel deep ensemble model that combines multiple machine learning algorithms, outperforming individual models and standard ensemble techniques in yield forecasting tasks.
3. We conduct an extensive comparative evaluation against various baseline and state-of-the-art models, demonstrating consistent improvements across multiple performance metrics (MAE, RMSE, R^2).
4. We analyse the generalisation capability of RicEns-Net by evaluating the train–test performance gap, confirming its robustness and reduced risk of overfitting compared to other models.
5. We validate the model on noisy, multi-source real-world data from the Mekong Delta, Vietnam, highlighting its practical utility and operational potential for precision agriculture applications.

The remainder of this paper is organised as follows: Section 2 reviews related work, analysing previous findings and establishing the foundation for the study's objectives and contributions. Section 3 details the materials and methods, including descriptions of data sources, preprocessing techniques, and feature engineering approaches. Section 4 presents the experimental results derived from the proposed model, followed by Section 5, which interprets these findings and discusses their implications. Finally, Section 6 summarises the study, outlines its limitations, and highlights potential directions for future research.

2. Related works

In parallel with rapid technological advancements over the past two decades, recent breakthroughs in machine learning (ML), artificial intelligence (AI), and remote sensing (RS) have dramatically expanded the capabilities and scope of research, particularly in environmental monitoring. These advancements have enabled groundbreaking achievements such as capturing the first-ever image of a black hole (Akiyama et al., 2019) and detecting floating marine plastic debris from space (Biermann et al., 2020; Booth et al., 2023). In addition, these technological innovations have significantly improved various environmental applications, ranging from air pollution monitoring (Rowley and Karakuş, 2023) and precise road detection (Ma et al., 2024; Akhtarmanesh et al., 2023) to comprehensive land cover mapping (Ma et al., 2022) and many others (Roy et al., 2023; Marzvan et al., 2021; Esmaeili et al., 2023; Safari et al., 2024). Driven by these developments, this section first outlines the specific objectives in Section 2.1, and subsequently provides a detailed review of key academic studies in Section 2.2 that highlight recent advances from the perspectives of crop yield prediction and agricultural monitoring, emphasising the transformative impact of ML, AI, and multi-modal RS data on agricultural decision-making and management practices.

2.1. Objectives

The primary objectives of this research are as follows:

- To integrate diverse remote sensing (SAR, optical imagery) and meteorological data sources into a coherent multi-modal dataset to enhance the accuracy and robustness of crop yield prediction.
- To address feature complexity by applying advanced feature selection techniques, thereby reducing dimensionality while preserving essential predictive information.

- To design a novel deep ensemble learning framework (RicEns-Net) capable of leveraging the strengths of multiple machine learning algorithms for improved yield forecasting.
- To extensively benchmark the proposed model against state-of-the-art machine learning techniques, demonstrating its superior performance using standard evaluation metrics (MAE, RMSE, R^2).
- To validate the proposed methodology on noisy, real-world multi-source data from the Mekong Delta region and assess its practical utility for operational deployment in precision agriculture.

2.2. Literature review

Advancements have influenced the evolution of crop yield monitoring in sensor technology and data analysis methodologies. While early studies highlighted rainfall as the primary determinant of crop yield, a key shift occurred in 1968 with the recognition of soil moisture as a more reliable predictor by Baier and Robertson (1968). Their work leverages spectral data to estimate crop yield based on vegetation health indicators. Over time, numerous vegetation indices (VIs) have been developed to assess vegetative conditions and physiological characteristics of crops. These VIs, including the Normalised Difference Vegetation Index (NDVI), Leaf Area Index (LAI) (Bouman, 1995), and Transformed Soil Adjusted Vegetation Index (TSAVI) (Baret and Guyot, 1991), play a crucial role in crop prediction models. Advancements in hyperspectral imaging have enabled the capture of fine-grained spectral data, facilitating the development of biochemical indices for quantifying plant constituents.

Rice crop yield estimation relies on understanding the crop's growth stages and environmental factors. Water level in paddy fields, rather than direct precipitation, is crucial for irrigated rice fields. Accumulating temperature is more important than temperature at certain times, as it affects the crop's development stages. These factors are integrated into crop models to predict yield accurately.

In the early 2000s, research surged, leveraging imaging and machine learning technologies for crop yield prediction. Studies introduced novel methodologies, such as artificial neural networks (ANN) and SVR, to analyse remote sensing data and historical yield records. These approaches demonstrated more precise results compared to traditional models and prepared for more precise and scalable methods for estimating crop yields. Uno et al. (2005) analyse hyperspectral images of corn plots in Canada using statistical and ANN approaches, demonstrating the potential of ANNs in predicting yield with higher accuracy compared to conventional models. Li et al. (2007) introduce a methodology employing ANN models to predict corn and soybean yields in the United States "corn belt" region, achieving high prediction accuracy through historical yield data and NDVI time series. Bala and Islam (2009) estimate potato yields in Bangladesh using TERRA MODIS reflectance data and Vegetation Indices (VIs), demonstrating the effectiveness of VIs derived from remote sensing for early yield estimation. Li et al. (2009) employ SVR and multi-temporal Landsat TM NDVIs to predict winter wheat yield in China, showcasing the precision and effectiveness of SVR models in yield estimation. Stojanova et al. (2010) integrate LiDAR and Landsat satellite data using machine learning techniques to model vegetation characteristics in Slovenia. Their approach combines the precision of LiDAR data with the broad coverage of satellite data, facilitating effective forest management and monitoring processes.

Furthermore, Mosleh et al. (2015) evaluated the efficacy of remote sensing imagery in mapping rice areas and forecasting production, highlighting challenges such as spatial resolution limitations and issues with radar imagery. Johnson et al. (2016) developed crop yield forecasting models for the Canadian Prairies, revealing the effectiveness of satellite-derived vegetation indices, particularly NDVI, in predicting yield potential. Pantazi et al. (2016) proposed a model for winter wheat yield prediction, integrating soil spectroscopy and remote sensing data to visually depict yield-influencing factors. Ramos et al. (2020) introduced an optimised Random Forest algorithm for maize-crop yield prediction, emphasising the importance of vegetation indices like NDVI, NDRE, and GNDVI. Li et al. (2021) utilised extreme gradient boosting machine learning to accurately predict vegetation growth in China, achieving high predictive accuracy and demonstrating effectiveness under diverse conditions. Zhang et al. (2021) employed field-surveyed data to predict smallholder maize yield, with novel insights into the performance of various vegetation indices and machine learning techniques.

Recent studies have demonstrated the effectiveness of using Sentinel-2 satellite imagery and machine learning techniques for predicting crop yields and mapping crop types. Son et al. (2022) employed Sentinel-2 image composites and various machine learning algorithms to forecast rice crop yields in Taiwan. Their results showed that Support Vector Machines (SVM) outperformed RF and ANN at the field level. This highlighted SVM's potential for accurate yield predictions approximately one month prior to harvesting. Perich et al. (2023) utilised Sentinel-2 imagery to map crop yields at the pixel level in small-scale agriculture. Their machine learning models, based on spectral indices and raw reflectance data, proved effective despite challenges posed by cloudy satellite image time series. Khan et al. (2023) combined ground-based surveys, Sentinel-2 satellite imagery, and deep learning techniques to map crop types. They successfully achieved high accuracy in identifying staple crops such as rice, wheat, and sugarcane within the first four weeks after sowing.

Along with Sentinel 2, UAV and other sensor spectral information have also been used in the literature in the last couple of years. Shafi et al. (2023) propose XGBoost, LASSO, and RF regression models to be utilised via Drone-based multispectral imagery. Islam et al. (2023) combine remote sensing and meteorological data in stacking multiple regression models for rice crop yield prediction. Zhou et al. (2023) compare CNN and LSTM-based models for predicting annual rice yield in Hubei Province, China. By utilising ERA5 temperature data and MODIS vegetation indices, they demonstrate that the CNN-LSTM model with spatial heterogeneity outperforms models using only remote sensing data. Arshad et al. (2023) evaluate the performance of RF and SVR, in predicting wheat yield in southern Pakistan using a combination of remote sensing indices and climatic variables, where RF outperforms other methods. Asadollah et al. (2024) assess the effectiveness of using a novel Randomised Search cross-validation (RScv) optimisation algorithm with four machine learning models. With a target to predict annual yields of four crops (Barley, Oats,

Rye, and Wheat) across 20 European countries, they demonstrate improved prediction accuracy through satellite-based climate and soil data.

Furthermore, Lu et al. (2024) present a state-of-the-art CNN-BiGRU model enhanced by GOA and a novel attention mechanism (GCBA) for accurate county-level soybean yield estimation in the U.S. They leverage multi-source remote sensing data and outperform existing models in yield prediction accuracy. Killeen et al. (2024) investigate UAV-based corn yield prediction using RF and linear regression models. Their findings show that spatial cross-validation reduces over-optimism in yield prediction compared to standard 10-fold cross-validation, with LR showing better spatial generalisability than RF. Dhaliwal and Williams II (Dhaliwal and Williams, 2024) use a 26-year dataset on US sweet corn production to evaluate machine learning models for yield prediction. The authors report that RF performs best, with year, location, and seed source identified as the most influential variables.

Recently, Gadupudi et al. (2024) demonstrated integrating ML strategies like RF and Decision Trees alongside DL models such as LSTM and RNN. To improve crop prediction accuracy, the authors incorporate soil attributes, climate data, and cost analyses to optimise results. Similarly, Rao et al. (2024) employed attention-based CNNs and bidirectional LSTMs with hyperparameter tuning to predict crop yields. They showcase significant improvements in detection performance through methods like the shuffling shepherd optimisation algorithm. Sharma et al. (2025) explored the fusion of AI algorithms, including logistic regression and IoT-enabled analytics, to tailor recommendations based on regional agricultural parameters, advancing productivity and diversification.

The diverse literature outcomes examined previously, along with numerous others, have highlighted the multidimensional potential of AI in transforming traditional agricultural practices into more data-driven, adaptive systems. They have employed a variety of data types from different sources and machine learning models. This diversity presents challenges in generalising techniques across different datasets, yet it also enhances performance for specific datasets. The adoption of multi-modal data usage, multi-modal AI techniques, and Ensemble methods has emerged as the current practice in this research field. Shahhosseini et al. (2021) explore the predictive performance of two novel CNN-DNN machine learning ensemble models for forecasting county-level corn yields across the US Corn Belt. The study compares the effectiveness of homogeneous and heterogeneous ensemble creation methods by combining management, environmental, and historical yield data from 1980 to 2019. The authors report that homogeneous ensembles provide the most accurate yield predictions, offering the potential for the development of a reliable tool to aid agronomic decision-making. Gavahi et al. (2021) introduce DeepYield, a novel approach for crop yield forecasting that combines Convolutional Long Short-Term Memory (ConvLSTM) and 3-Dimensional CNN (3D-CNN). By integrating spatiotemporal features extracted from remote sensing data, including MODIS Land Surface Temperature (LST), Surface Reflectance (SR), and Land Cover (LC), DeepYield outperforms traditional methods and demonstrates more precise forecasting accuracy for soybean yields across the Contiguous United States (CONUS). Zare et al. (2024) investigate the impact of data assimilation techniques on improving crop yield predictions by assimilating LAI data into three single crop models and their multimode ensemble using a particle filtering algorithm. The results of their case study in southwestern Germany reveal that data assimilation significantly enhances the accuracy of the LAI simulation and the prediction of grain yield. Particularly for certain crop models, they highlight the potential for further improvements in data assimilation applications through regional model calibration and input uncertainty analysis.

Recent studies highlight the effectiveness of ensemble learning techniques in enhancing the accuracy and reliability of crop yield predictions. Gopi and Karthikeyan (2024) proposed the Red Fox Optimisation with Ensemble Recurrent Neural Network for Crop Recommendation and Yield Prediction (RFOERNN-CRYP) model. Their model leverages an ensemble of deep learning architectures, including LSTM, bidirectional LSTM (BiLSTM), and gated recurrent units (GRU), optimised through the Red Fox Optimisation algorithm. This approach demonstrated superior performance compared to individual classifiers, providing automated crop recommendations and accurate yield predictions to assist farmers in decision-making. Similarly, Boppudi et al. (2024) developed a deep ensemble model integrating Deep Max Out, Bi-GRU, and CNN architectures, along with advanced preprocessing and feature selection using the IBS-BOA algorithm. Their method effectively addressed variability in weather and environmental conditions, significantly reducing prediction error rates compared to existing approaches.

Additionally, stacking and weighted ensemble methods have been successfully employed in recent research. Umamaheswari and Madhumathi (2024) utilised a stacking ensemble comprising SVR, KNN, and RF as base learners, with LASSO regression as the meta-learner, achieving enhanced precision in crop yield predictions. Osibo et al. (2024) integrated weighted ensemble techniques with remote sensing data, outperforming state-of-the-art models while simplifying the data integration process. Furthermore, Zhang et al. (2024) introduced the StackReg framework, combining UAV-acquired multispectral data with ridge regression, SVM, Cubist, and XGBoost. Their ensemble consistently achieved better performance than individual base models, particularly in multi-stage prediction scenarios. Collectively, these studies underscore the critical role of ensemble learning in agricultural applications, offering robust, adaptable solutions across various agricultural contexts.

Furthermore, Mahdipour et al. (2024) proposed an “ultrafusion” framework that models inherent uncertainty in multiple high-resolution panchromatic images using fuzzy number theory and fusion-based segmentation. Their approach effectively captures the randomness introduced during image formation and preprocessing, significantly improving segmentation metrics such as overall accuracy and F1-score. Meanwhile, Mirhoseini Nejad et al. (2024) introduced an integrated framework combining 3D Convolutional Neural Networks (3D-CNNs), ConvLSTM, and Vision Transformers (ViTs) for soybean yield prediction using multispectral remote sensing data. By leveraging spatial, temporal, and global contextual features simultaneously, their hybrid model substantially outperformed existing methods in yield prediction tasks, highlighting the importance of combining multiple learning architectures for complex agricultural datasets. Furthermore, Farmonov et al. (2024) developed HypsLiDNet, a deep learning model designed for hyperspectral imaging (HSI) and LiDAR data fusion in crop classification. By incorporating morphological operations and attention mechanisms, HypsLiDNet achieved notable improvements in classification accuracy compared to both traditional machine learning and contemporary deep learning techniques. These works collectively underline the emerging trend of integrating multi-modal

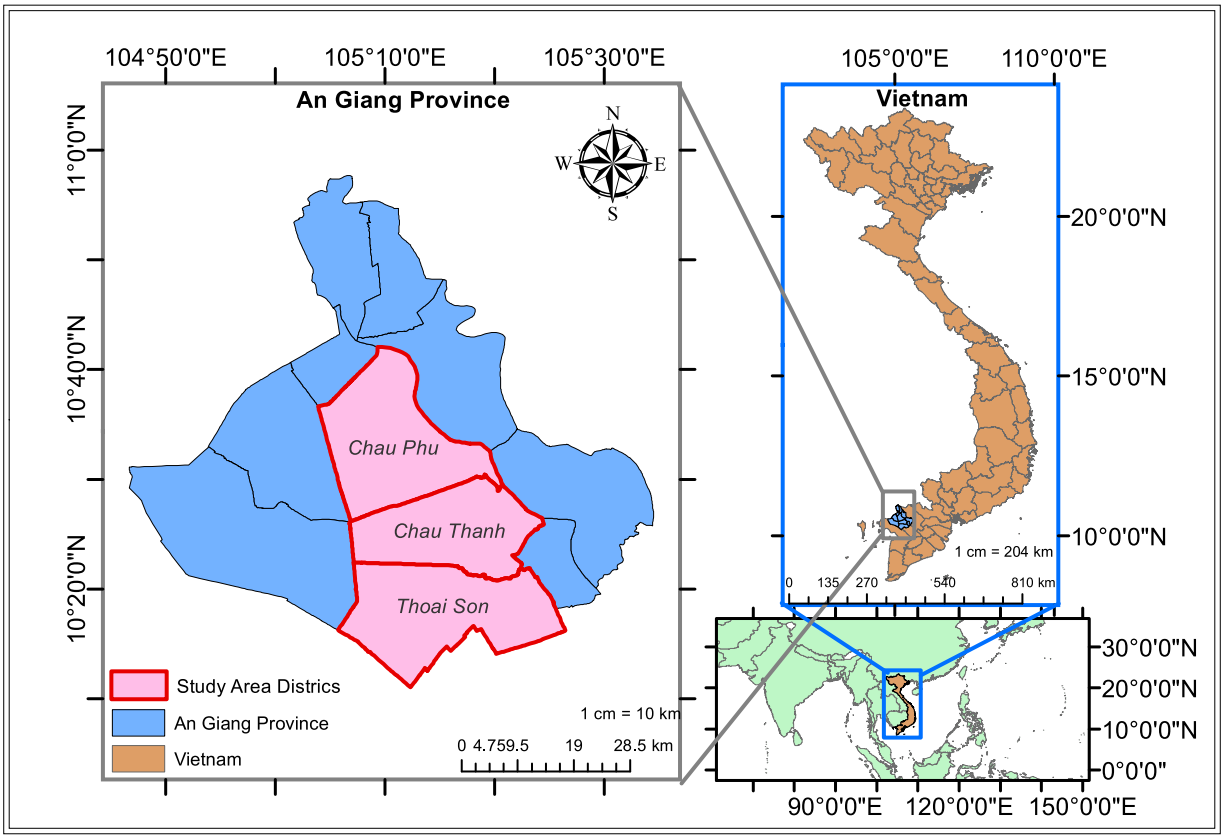

Fig. 1. Map illustrating the geographical area under investigation, encompassing the An Giang province in Vietnam, including the specific study districts of Chau Thanh, Thoai Son, and Chau Phu. This region is significant for its rice production and diverse environmental conditions, making it an ideal case study for testing the generalisability of the RicEns-Net model.

remote sensing data and advanced ensemble learning techniques to tackle the challenges of agricultural monitoring and prediction with greater robustness and precision.

Building upon the ensemble learning advancements discussed above, it is evident that integrating diverse multi-modal data and advanced ensemble architectures significantly improves crop yield predictions. However, despite substantial progress, existing methods typically focus on a limited set of data sources and conventional deep learning architectures. To address these gaps, we introduce a novel framework for predicting crop yields, named "*RicEns-Net*". This framework incorporates advanced data engineering processes involving five unique data sources, namely Sentinel 1/2/3, NASA's Goddard Earth Sciences (GES) Data and Information Services Centre (DISC), and field measurements. The novelty of RicEns-Net lies in its sophisticated integration of these diverse and rich multi-modal datasets, comprising 15 carefully selected features from an initial pool exceeding 100. Additionally, RicEns-Net employs an advanced Deep Ensemble approach, combining widely-used architectures such as CNN and MLP with less frequently utilised DenseNet and Transformer structures, thus addressing gaps and expanding upon the current state-of-the-art.

3. Materials and methods

3.1. Study area & rice crop details

As stated earlier, this study begins by employing the dataset offered by EY for the 2023 iteration of their Open Science Data Challenge (EY, 2023). The dataset encompasses information from 557 farm sites situated in Chau Thanh, Thoai Son, and Chau Phu districts within the province of An Giang in Vietnam (see Fig. 1). The study province of An Giang relies significantly on agriculture as a cornerstone of its economy. Notably, An Giang province is situated in the Mekong River delta region, crucial for providing irrigation to support rice cultivation. The dataset, supplied by EY, contains fundamental details for each crop, including District Name, Latitude, Longitude, Crop Season, Crop Frequency, Harvest Date, Crop Area, and Yield as given in Table 1.

Every entry in the primary dataset represents an individual crop and is characterised by eight features, including three categorical variables (District name, Crop Season [WS = Winter Spring; SA = Summer Autumn], and Crop Frequency of the specified farm [D = Twice; *T* = Thrice]) and five numerical variables (Latitude, Longitude, Harvest Date, Area [Hectares], and Yield Rate [kg/Ha]). The

Table 1
Study area details, including district names, geographic locations, population, area, and population density for Chau Thanh, Thoai Son, and Chau Phu districts in An Giang province, Vietnam. This information provides context for the study region's significance in rice production.

District	Province	Data count	Geographic location	Population	Area density	Population
Chau Thanh	An Giang	218	Mekong Delta Region	130,101	571 km^2	228/km^2
Thoai Son		171		187,620	456 km^2	411/km^2
Chau Phu		168		250,567	426 km^2	588/km^2

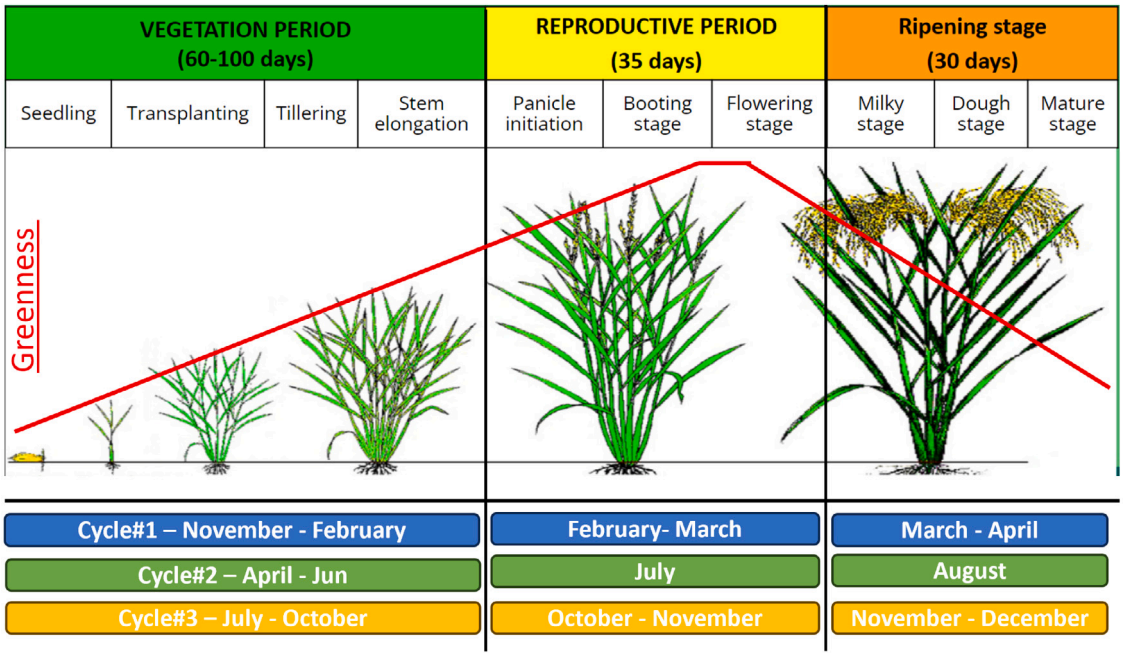

Fig. 2. Rice growing stages and three potential crop cycles in Vietnam's study region. Credit: (e-Extension, 2024). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

harvesting dates of the crops in question extend from March 18, 2022, to August 9, 2022, covering two significant crop seasons, namely Summer–Autumn and Winter–Spring.

While the overall duration of the crop cycle spans approximately 5–6 months, contingent on the season, the period from transplanting to harvesting typically ranges from 90 to 130 days. The growth trajectory can be categorised into three principal phases: Vegetative, Reproductive, and Ripening stages, illustrated in Fig. 2. During the Reproductive stage, rice plants attain their maximum greenness, marking the culmination of this phase. Subsequently, the crop transitions into the Ripening stage, characterised by the transformation of the plants from green to yellow, coinciding with the maturation of rice grains. In the context of Vietnam, rice cultivation occurs biannually or triannually within two seasons: Winter–Spring (Nov–Apr), Summer–Autumn (Apr–Aug), and Autumn–Winter (Jul–Dec).

The necessary geospatial information can be obtained from either Landsat or Sentinel satellites, taking into account the designated location and harvest time outlined in the dataset. After careful consideration of technical details such as Ground Spatial Resolution (GSD) and revisit frequency, the decision is made to opt for data sourced from Sentinel satellites. Additionally, there is anticipation in leveraging the complete spectrum of SAR, MSI, and Meteorological data. To gather remote sensing data for a specified area, it is essential to finalise two key parameters: the time window and the geographical bounding box or crop window.

We intend to gather data during the phase when the crop is nearing complete maturation. The progress of the crop's growth is discerned through the intensity of the green hue in the data, with the period from transplantation to maturity spanning 60–100 days, contingent on the prevailing season. Following full maturity, the crop progresses into the ripening stage, which typically lasts around 30 days. To determine the timeframe for data collection, we use the harvest date as a reference point. The initiation date is established as 60–90 days before the harvest date, while the conclusion date is set at 30 days before the harvest date (refer to Fig. 2). This designated time window encapsulates the entire duration from the crop's transplantation to the conclusion of the maturity stage, ensuring that the remote sensing data aligns with the various growth phases of the crop.

The primary input data exclusively provides information about the field's area without specifying the precise bounding region for each field. This lack of detail hinders our ability to extract data for the exact boundaries of individual crop fields. To address

this limitation, we proceed by identifying the precise location of each field in the MSI data and cropping the image to obtain a set of pixels surrounding that specific location. Given that the visual bands of the MSI data have a spatial resolution of 10 m per pixel, a 3 × 3 pixel collection corresponds to an area of 900 square meters, equivalent to 0.222 acres or 0.09 ha. Notably, the SWIR16 and SWIR22 bands of Sentinel 2's MSI exhibit a spatial resolution of 20 m per pixel, meaning a 3 × 3 pixel bounding box represents an area of 3600 square meters, or 0.89 acres, or 0.36 ha.

3.2. Data/feature extraction

3.2.1. Sentinel-1

Sentinel-1 operates with four distinct acquisition modes: Stripmap (SM), Interferometric Wide Swath (IW), Extra Wide Swath (EW), and Wave (WV). During Synthetic Aperture Radar (SAR) acquisition, known as the "datatake" process, data from SM, IW, and EW modes is divided into smaller, manageable slices. These slices undergo processing to generate different product levels. Level-0 products contain raw sensor data, while Level-1 products provide calibrated imaging data such as Single Look Complex (SLC) and Ground Range Detected (GRD) formats, essential for interferometry and georeferenced analysis. Finally, Level-2 products include higher-level geophysical measurements, such as Ocean (OCN) products, which support oceanographic and environmental applications.

Sentinel 1 delivers SAR images featuring two polarisations, vertical (VV) and horizontal (VH), characterised by the difference in the polarisation of their transmitted and received signals. The polarisation of radar signals plays a crucial role in deciphering the structure and orientation of surface elements on the land. The radar signal experiences scattering and depolarisation due to the randomly oriented structure of plant leaves as it undergoes multiple bounces. By comparing the vertical (VV) and horizontal (VH) components, the degree of scattering by the land surface can be discerned. In our model, we incorporate this technique through the VV/VH_{ratio} feature. The data were collected by defining a geographical bounding box with a size of 0.001 degrees, resulting in an output array size of 3 × 3. However, some location data did not conform to this shape and required trimming to achieve a 3 × 3 box size. Given Sentinel 1's spatial resolution of 10 m, each box corresponds to an area of 30 m by 30 m. The subsequent features were derived from Sentinel 1 SAR data

- Set 1 (4 variables) → $VV_{mean}, VH_{mean}, VV/VH_{ratio,mean}, RVI_{mean}$

where the radar vegetation index (RVI) is given as

$$RVI = \frac{VV}{VV + VH}. \tag{1}$$

3.2.2. Sentinel-2

Sentinel-2 imagery is collected through a continuous acquisition process known as 'datatake', covering up to 15,000 km in length. The acquired data are structured into products at different processing levels. Sentinel-2 products are organised into tiles or granules, following a Universal Transverse Mercator (UTM) projection system to ensure global coverage.

At Level-1B, the data are provided in granules of 23 km × 25 km, containing radiometrically calibrated and geometrically refined imagery. Level-1C and Level-2A products are structured into 110 km × 110 km tiles, ensuring seamless coverage across designated UTM zones. Each tile overlaps with neighbouring ones to maintain spatial continuity and facilitate multi-temporal analysis. Level-1C includes top-of-atmosphere (TOA) reflectance, while Level-2A applies atmospheric corrections to generate bottom-of-atmosphere (BOA) reflectance, making it suitable for land surface monitoring and vegetation analysis.

Sentinel 2 data furnish spectral intensities across 13 bands, encompassing Visual-NIR (VNIR) to Short-Wave Infra-Red (SWIR) regions. Notably, there are four spectral bands, namely Red, Green, Blue, and NIR (B04, B03, B02, and B08), offering a ground resolution of 10 m. Additionally, six bands exhibit a 20-m ground resolution, comprising four Red Edge bands (B05, B06, B07, and B08A) and two SWIR bands with distinct wavelengths (B11 and B12). The remaining three bands, with a 60-m ground resolution, serve specific purposes: B01 for aerosol detection (0.443 μm), B09 for water vapour observation (0.945 μm), and B10 for cirrus detection (1.374 μm). Notably, the Sentinel 2 mission boasts a revisit frequency of 5 days.

When acquiring MSI data, it is crucial to account for the potential impact of cloud cover in the targeted area. With a revisit frequency of 5 days, there are only 6–8 chances to capture images during the period when crops reach full growth before maturation. Given Vietnam's tropical monsoon climate, these image opportunities are prone to cloudiness. The dataset at hand reveals median cloud coverage values of 16% and 21% during the Winter–Spring and Summer–Autumn seasons, respectively.

In order to prevent the occurrence of unclear or cloudy images, it is necessary to eliminate those with a high level of cloud coverage. Simultaneously, we aim to capture comprehensive crop data when the plants are at their full growth and exhibit maximum greenness. To achieve optimal outcomes, we conducted experiments with various values for maximum cloud coverage and time windows. Based on the findings detailed in Table 2, we concluded that setting the maximum cloud coverage to 60% and collecting Sentinel-2 MSI data during the 50 days preceding the crop maturation provides favourable results. The objective is to secure a minimum of 4–5 images for each specific location.

To ensure that the spectral intensity trends are captured, we identify the minimum, maximum, mean and variance of 9 MSI bands based on all the MSI images available for each location.

- Set 2 (36 variables) → is in a format of "$Band_{stats}$" where $Band = \{B02, B03, B04, B05, B06, B07, B08, B11, B12\}$ and $stats = \{min, max, mean, var\}$.

Table 2
Trials for identifying optimal value of cloud coverage threshold & time window for data collection.

Trial	Max. coverage	Window (*pre*-Maturity)	No. of crops							Remarks
			0	1	2	3	4	5	>5	
1	25%	30	119	298	138	0	2	0	0	**Reject**
2	30%	30	119	298	138	0	0	2	0	**Reject**
3	40%	30	13	348	83	109	2	0	2	**Reject**
4	25%	40	5	412	138	0	2	0	0	**Reject**
5	25%	45	5	380	170	0	2	0	0	**Reject**
6	25%	50	5	358	114	78	0	2	0	**Reject**
7	30%	50	5	252	204	94	0	0	2	**Reject**
8	40%	50	0	250	204	94	0	2	2	**Reject**
9	50%	50	0	0	57	193	0	15	287	**Reject**
10	**60%**	50	0	0	0	0	70	200	287	**Accept**

MSI data has been used to create transformational features known as *Vegetation Indices* given in Table A.5 like NDVI, SR, EVI, EVI2, SAVI, RGVI, DVI, MSR, NIRv, kNDVI, NDVIre, NDRE1, NDRE2 to indicate the volume of vegetation on the land surface.

- Set 3 (26 variables) → is in a format of "VI_{stats2}" where $VI = \{NDVI, SR, EVI, EVI2, SAVI, RGVI, DVI, MSR, NIRv, kNDVI, NDVIre, NDRE1, NDRE2\}$ and $stats2 = \{mean, var\}$.

Utilising MSI data, various features, such as NDWI, BSI, and LSWI, as outlined in Table A.5, have been generated to depict soil and water content. This application is particularly advantageous in the context of rice cultivation, where the crop is submerged in water.

- Set 4 (6 variables) → is in a format of "$VI2_{stats2}$" where $VI2 = \{NDWI, BSI, LSWI\}$ and $stats2 = \{mean, var\}$.

Additional features have been generated using MSI spectral data, incorporating information derived from the biochemical properties of the plants (refer to Table A.5).

- Set 5 (4 variables) → is in a format of "$VI3_{stats2}$" where $VI3 = \{CCI, GCC\}$ and $stats2 = \{mean, var\}$.

3.2.3. Sentinel-3

Sentinel-3 ensures the continuous availability of high-quality data for monitoring land, ocean, and atmospheric conditions, particularly in coastal areas where accuracy is critical. The mission provides comprehensive environmental observations globally, supporting a range of applications. Sentinel-3 plays a key role in fire detection, inland water surface height measurements, and land ice/snow surface temperature assessments. Its multi-instrument payload enables precise monitoring of ocean colour, sea surface temperature, and land surface dynamics, contributing to climate research, water resource management, and disaster response.

Sentinel 3 data were acquired to obtain meteorological information related to environmental variables such as ambient air temperature (MET_{temp}), land surface temperature (LST), solar radiation (MET_{solrad}), and specific humidity (MET_{sh}). These data sets have been integrated into the model as the following features:

- Set 6 (8 variables) → is in a format of "$S3_{stats2}$" where $S3 = \{MET_{temp}, LST, MET_{solrad}, MET_{sh}\}$ and $stats2 = \{mean, var\}$.

3.2.4. NASA GES DISC

Rainfall information was acquired from NASA's Goddard Earth Sciences (GES) Data and Information Services Centre (DISC) through the utilisation of the Google Earth Engine API. The data retrieval involved the utilisation of the *precipitationCal* parameter, which denotes rainfall in mm per hour. We organise this data into two distinct features: $\text{Rainfall-Total}_{growth}$ and $\text{Rainfall-Total}_{maturity}$, encompassing three statistical measures—mean, maximum, and sum as

- Set 7 (6 variables) → is in a format of "$NASAGES_{stats3}$" where $NASAGES = \{Rainfall_{growth}, Rainfall_{maturity}\}$ and $stats3 = \{mean, max, sum\}$.

3.3. Correlation analysis

After meticulously gathering all potentially valuable engineered features from multi-modal remote sensing data, we proceed to examine their statistical and predictive capabilities for subsequent feature selection. This phase, delineated in this sub-section, initiates with a correlation analysis.

Concerning the relationship between SAR features and $Yield_{rate}$, all four data features exhibit a strong correlation with $Yield_{rate}$. As anticipated, the VH_{mean} feature effectively captures the backscattering of the SAR signal by the rice plant leaves, resulting in a higher correlation (0.32) with $Yield_{rate}$ compared to VV_{mean} (0.25). The VV/VH_{ratio} serves as a transformative feature, demonstrating an enhanced correlation (0.45) in comparison to both VV_{mean} and VH_{mean} individually. Notably, the Radar Vegetation Index (RVI) shows a similar positive correlation (0.41) with the Yield Rate.

Table 3
Engineered and extracted features after selection stages.

Variable	Description	Type	Source
$\text{Season}_{\text{Enc}}$	Crop season indicator	Categorical	Field
$\text{Dist}_{\text{Chau Phu}}$	Crop location indicator	Categorical	Field
$\text{Dist}_{\text{Chau Thanh}}$	Crop location indicator	Categorical	Field
$\text{Dist}_{\text{Thoai Son}}$	Crop location indicator	Categorical	Field
$Yield_{kg}$	Rice yield in kg at a specific point	Numerical	Field
$Rainfall_{growth,max}$	Max rainfall growth in mm per hour	Numerical	NASA GES DISC
$Rainfall_{growth,sum}$	Sum rainfall growth in mm per hour	Numerical	NASA GES DISC
$Rainfall_{maturity,max}$	Max rainfall maturity in mm per hour	Numerical	NASA GES DISC
VV_{mean}	Mean SAR image intensity in VV polarisation	Numerical	Sentinel-1
$B08_{max}$	Max NIR spectral band intensity	Numerical	Sentinel-2
$RGVI_{mean}$	Mean spectral index	Numerical	Sentinel-2
$kNDVI_{mean}$	Mean spectral index	Numerical	Sentinel-2
GCC_{mean}	Mean spectral index	Numerical	Sentinel-2
LST_{mean}	Mean land surface temperature	Numerical	Sentinel-3
$MET_{solrad,mean}$	Mean solar radiation	Numerical	Sentinel-3
$Yield_{rate}$	**Rice yield in kg per hectare** (kg/ha)	**Target**	**Field**

As previously indicated regarding Sentinel-2 data, we derive spectral statistics from 9 bands: B02 Blue, B03 Green, B04 Red, B05-B07 Red Edge, B08 NIR, and B11-B12 SWIR. These statistics, namely min, max, mean, and variance, are incorporated into the model. Upon analysing observations across all bands, it is noteworthy that variance features exhibit a relatively low correlation (<0.1) with the target variable, whereas other statistical measures demonstrate correlation coefficient values surpassing 0.3.

Spectral data are employed to generate transformative characteristics known as Vegetative Indices (VI). These indices serve as a more efficient measure for discerning and monitoring variations in plant phenology. In our approach, we utilise VIs such as NDVI, NDVIre, NDRE1, NDRE2, SR, DVI, MSR, EVI, EVI2, SAVI and RGVI, NIRv, and kNDVI. These features are integrated into the model in the form of their respective mean and variance features. However, the variance feature is omitted from the model due to its limited correlation with the target variable. Notably, kNDVI exhibits one of the highest correlations with Yield, while features like DVI, EVI, and NIRv demonstrate some of the lowest correlations.

Similar to the vegetation indices, we can employ optical data to compute additional indices that precisely quantify the environmental conditions of the crop's cultivation. NDWI, LSWI16, and LSWI20 specifically indicate the water or moisture content in the soil, which is crucial for rice cultivation, requiring flooded fields. Conversely, BSI reflects the soil condition. We incorporate these attributes into the model as their respective Mean and Variance features. Similar to the approach with vegetation indices, we have excluded the variance feature from the models due to its limited correlation with Yield. Notably, all water indices exhibit high correlations with each other and share a similar correlation with Yield.

Lastly, concerning meteorological characteristics, the average ambient air temperature (refers to variable "$MET_{temp,mean}$") and specific humidity ($MET_{sh,mean}$) exhibit the strongest correlation with crop yield, and they also demonstrate a high-degree of correlation between themselves. Solar radiation ($MET_{solrad,mean}$) emerges as a significant predictor due to its notable correlation with yield and comparatively lower correlation with other meteorological features.

3.4. Feature selection

Up to this point, all the engineered features, totalling 94 in number, have undergone various stages of processing. These stages include (i) grouping, involving the arrangement of data types and condensation into categorical, numerical, and object types; (ii) scaling, which entails MinMax scaling; and (iii) splitting through a train–test split with a ratio of 3:1.

As outlined in the preceding sections, the subset of the 94 features exhibits significant correlation, and incorporating all these features in the models would lead to computationally intensive experiments. To mitigate this, during the final processing stage, we execute multiple rounds of feature selection, including Pairwise Feature Independence Check using the χ^2 test, statistical significance tests based on p-values, outlier removal, and thresholding for correlation and variance. Following these stages, the outcome is a refined set of 15 predictive (11 numerical & 4 categorical) features and 1 target feature all of which are shown in Table 3.

The whole data collection, processing and engineering stages are plotted in Fig. 3 whilst a detailed breakdown block diagram of the Data Engineering - 2 is also presented in Fig. 4.

3.5. The proposed model - RicEns-Net

After the extensive stages of data engineering, pre-processing, and preparation mentioned earlier, the 15 most informative and significant data features from 5 different data modalities will be employed to predict rice yield for the specified locations. This section provides a comprehensive introduction to the proposed deep ensemble model, *RicEns-Net*. As mentioned in the earlier stages, deep learning models have currently been dominating yield prediction studies. This paper stands parallel with these advances in the literature but tries to explore complementary advantages of different deep learning regression techniques under a deep ensemble architecture named RicEns-Net. The details of the RicEns-Net model architectures are presented in Fig. 5.

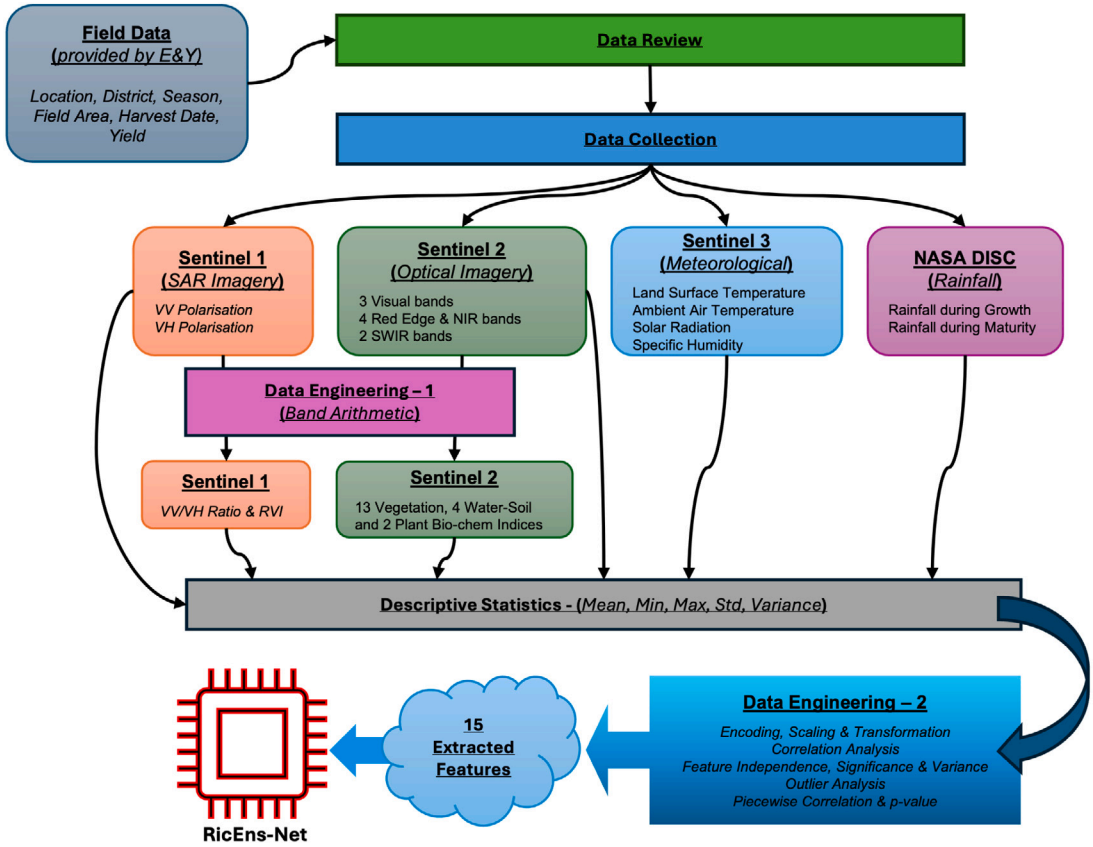

Fig. 3. Data collection, processing and engineering stages (Please see Figs. 4 and 5 for the details of the Data Engineering-2 block and RicEns-Net model, respectively.).

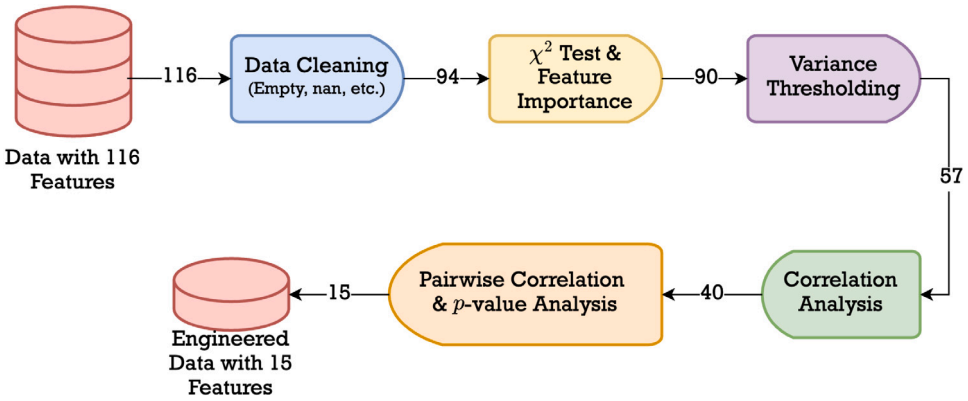

Fig. 4. A breakdown of the Data Engineering - 2 block.

RicEns-Net utilises two foundational deep learning architectures, CNN and MLP. Although various adaptations of these models have been proposed in the literature, no universally optimal architecture has emerged for crop yield prediction, prompting us to tailor our own implementations for this specific task. In addition to CNN and MLP, RicEns-Net incorporates four other powerful architectures: DenseNet, AutoEncoder (AE), U-Net, and TabTransformer. While these models are relatively underexplored in yield prediction tasks, they are prominent in related fields such as remote sensing, structured tabular learning, and image segmentation, offering complementary strengths to our ensemble.

Our motivation for including these diverse architectures lies in their unique contributions to learning from complex, multi-source agricultural data. AE architectures enable efficient feature compression by learning lower-dimensional representations that

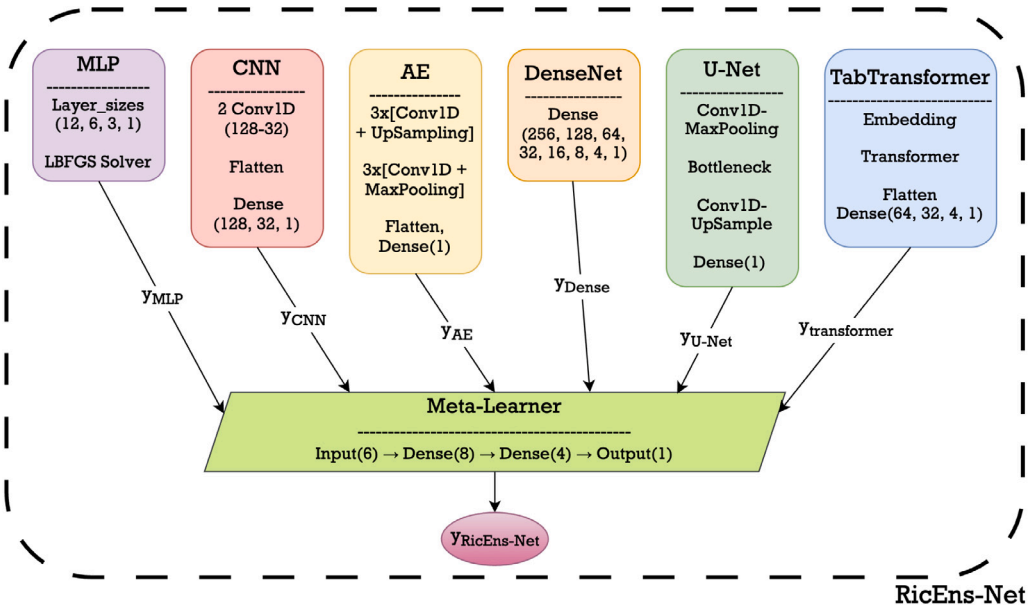

Fig. 5. RicEns-Net model details.

retain the most relevant information while mitigating noise, a valuable trait when handling noisy remote sensing data. DenseNet facilitates strong gradient flow and reuse of features through dense connectivity, making it especially effective in learning subtle and hierarchical patterns. U-Net, a widely used architecture in segmentation tasks, introduces a symmetric encoder–decoder structure with skip connections, which enhances spatial feature localisation, a critical factor for geospatially-aware yield prediction. Finally, TabTransformer brings attention mechanisms into tabular learning by integrating embeddings with transformer layers, enabling the model to capture feature interactions and contextual dependencies that traditional models often overlook. The complementary nature of these architectures strengthens RicEns-Net's generalisation ability and predictive accuracy.

In developing the RicEns-Net ensemble model, extensive testing was conducted to determine the optimal architectures for each individual deep learning model. This rigorous process involved systematically evaluating various configurations, including the number of layers, types of activation functions, and loss functions. For each model, we experimented with different depths ranging from shallow to deep architectures to find the optimal balance between complexity and performance. We also tested various activation functions, such as ReLU, sigmoid, SiLU, and tanh, to identify the most effective function for each model. Additionally, we compared multiple loss functions, including mean squared error, mean absolute error, and Huber loss, to select the one that minimised prediction errors most effectively. Furthermore, each model's architecture was refined through a series of experiments, incorporating cross-validation and hyperparameter tuning, to ensure the best performance for predicting rice yield. This exhaustive testing process ensured that each model within the ensemble was optimally configured to contribute to the overall predictive power of RicEns-Net. The details of these architectures, including the final selected configurations, are illustrated in Fig. 5. These configurations were chosen based on comprehensive performance evaluations, ensuring that each component model enhances the proposed deep ensemble's accuracy and robustness.

To generate the ensemble output of RicEns-Net, we adopt a meta-learning strategy instead of a fixed weighted average. In this approach, individual predictions from the six constituent models: CNN, MLP, DenseNet, AutoEncoder (AE), U-Net, and TabTransformer are used as input to a shallow neural network, referred to as the *meta-learner*. This meta-learner is trained to learn an optimal, non-linear combination of the base model outputs, effectively capturing complex inter-model dependencies and residual patterns that static weighting schemes cannot model.

Let y_i represent the prediction from the ith base model, where $i \in \{\text{MLP, CNN, Dense, AE, U-Net, transformer}\}$. The input vector to the meta-learner is then defined as

$$\mathbf{y} = [y_{\text{MLP}}, y_{\text{CNN}}, y_{\text{Dense}}, y_{\text{AE}}, y_{\text{U-Net}}, y_{\text{transformer}}]^{\top}$$

. The meta-learner is a shallow feedforward neural network that takes $\mathbf{y}$ as input and outputs the final prediction $y_{\text{RicEns-Net}}$. The network is trained using a held-out validation set, where the true yield values are used to supervise the learning of the aggregation function. This allows the meta-learner to assign flexible, data-driven importance weights to each base model output.

This dynamic architecture enables RicEns-Net to adapt its fusion strategy to different input scenarios and error behaviours, rather than relying on heuristics or static assumptions. Empirically, this approach provides superior generalisation and lower predictive error compared to traditional ensemble methods based on fixed weightings.

3.6. On deciding state-of-the-art

Despite significant advancements in computer vision research through the development of novel deep learning architectures, yield prediction research predominantly relies on traditional deep learning methods like CNN (Gavahi et al., 2021; Boppudi et al., 2024; Hashemi et al., 2024) and MLP (Maimaitijiang et al., 2020; Wu et al., 2024; Hussain et al., 2024), as well as state-of-the-art machine learning algorithms such as RF (Shahhosseini et al., 2021; Abdali et al., 2024; Manafifard, 2024), XGBoost (Shahhosseini et al., 2020, 2021; Li et al., 2024), and SVR/SVM (Chen et al., 2022; Tao et al., 2023; Abdali et al., 2024). According to a recent (2024) systematic review in Trentin et al. (2024), nearly one-third of the papers published in this area propose one of RF, SVM/SVR or CNN as the best-performing models. Additionally, traditional time series deep learning techniques, including Long Short-Term Memory (LSTM) (Boppudi et al., 2024; Gopi and Karthikeyan, 2024) and its variants, have been extensively utilised in the literature. Given this context, selecting the appropriate state-of-the-art models for the comparison study in this paper is of paramount importance. This section is dedicated to explaining the rationale behind the choice of each comparison model.

As previously discussed, this paper utilises a dataset from the EY Open Science Challenge 2023. This global data science competition, sponsored by EY, a leading consulting firm, is notable for its innovative application of Data Science and Analytics to real-world business scenarios. The EY Open Science Challenge 2023 ran for two months, from January 31, 2023, to March 31, 2023, attracting over 13,000 participants and more than 7500 submissions. The competition awarded USD 10,000 to the winner and USD 5000 to the runner-up. EY has provided performance results for the winning teams (global and employees) on the same test dataset, making these two models ideal candidates for comparison in this study. The global winner used a CatBoost regression model (referred to as CatBoost-EY in this paper), while the employee winner's model was based on Extremely Randomised Trees (referred to as ExtRa-EY). We present these models' performance metrics without infringing on their copyrights.

Considering that our ensemble model incorporates six deep learning architectures, we also evaluate their individual performance in the comparison study. This approach aligns with the systematic review paper (Joshi et al., 2023), which notes that 78 out of 102 proposed models in the literature up to 2023 include these architectures. This validation supports our choice of comparison models, ensuring that the proposed RicEns-Net is benchmarked against state-of-the-art standards. Additionally, we incorporate advanced machine learning algorithms such as XGBoost, RF, SVR, AdaBoost, CatBoost, ElasticNet, and Gradient Boosting into the comparison pool, along with their Voting and Stacked ensemble models (Keerthana et al., 2021; Abdali et al., 2024).

3.7. Implementation of models

We conducted our research using Python version 3.10.9, leveraging the rich ecosystem of libraries available for data science and machine learning. Our desktop workstation, featuring a robust 20-core processor and ample 32 GB of RAM, provided the computational power necessary for handling large datasets and complex modelling tasks efficiently. The widely used *scikit-learn* (*sklearn*) module served as the cornerstone for implementing traditional and state-of-the-art machine learning algorithms, offering a comprehensive suite of tools for data preprocessing, model selection, and evaluation. Additionally, for the implementation of deep learning architectures, we employed the widespread Python libraries *tensorFlow* and *keras*, which provide powerful abstractions and efficient computation frameworks tailored specifically for neural network development. These Python modules enabled us to explore a diverse set of modelling techniques and methodologies, ultimately facilitating the realisation of our research objectives with precision and scalability.

The dataset was divided into training and test sets with a split ratio of 3:1. The training set comprised 75% of the data, while the remaining 25% was reserved for testing. To ensure robustness in the model evaluation process, a 10-fold cross-validation (CV) procedure was applied exclusively to the training data. During this stage, the training data were split into 10 subsets, with each subset used as a validation set once while the model was trained on the remaining 9 subsets. This process was repeated 10 times to account for variability in training, and the results were averaged to minimise bias and variance. The test data, which was kept entirely separate, was evaluated only once after the training phase to assess the final model performance. This approach ensures that the test set remains unbiased by the training process and provides a reliable estimate of the model's generalisation capability.

3.8. Evaluation metrics

All the models will undergo assessment utilising regression metrics such as RMSE (Root Mean Square Error), MAE (Mean Absolute Error), R^2 Score (Coefficient Of Determination), and Adjusted R^2 (Adjusted Coefficient Of Determination). While MAE stands as the most straightforward performance metric, RMSE poses a more rigorous criterion by squaring the prediction error before calculating its mean and taking the square root. MAE and RMSE exhibit differences in their sensitivity to outliers.

$$MAE = \frac{1}{N}\sum\left(y_{predicted} - y_{actual}\right) \tag{2}$$

$$RMSE = \sqrt{\frac{1}{N}\sum\left(y_{predicted} - y_{actual}\right)^2} \tag{3}$$

Another critical metric is R^2, which is anticipated to fall within the 0 to 1 range, although it can dip below 0 for specific models. In simpler terms, the R^2 value gauges the model's capacity to elucidate the variance of the target variable. To provide a more precise definition, as outlined in the SkLearn user guide, it signifies the proportion of variance (of y) explained by the independent variables

Table 4
Rice crop yield prediction performance for all the utilised models. Bold and underlined values belong to the best and second-best models, respectively.

Models	RMSE	MAE	Train R^2	Test R^2	10-fold CV Avg. R^2	R^2 Diff.	Test Adj. R^2
Bayesian ARD	453.788	366.626	0.692	0.603	0.678 ± 0.055	0.088	0.589
U-Net	459.883	369.199	0.702	0.593	0.644 ± 0.061	0.109	0.578
kNN	475.213	379.494	0.739	0.565	0.665 ± 0.062	0.174	0.550
TabTransformer	440.256	345.702	0.704	0.627	0.657 ± 0.049	0.077	0.614
RF	449.309	361.096	0.733	0.611	0.678 ± 0.059	0.122	0.597
GB	452.787	354.215	0.814	0.605	0.656 ± 0.057	0.209	0.591
SVR	448.902	345.711	0.689	0.612	0.670 ± 0.051	0.077	0.598
CatBoost	459.893	361.478	0.792	0.593	0.666 ± 0.059	0.199	0.578
ElasticNet	450.985	351.660	0.688	0.608	0.668 ± 0.058	0.080	0.594
Huber	454.450	359.668	0.695	0.602	0.668 ± 0.059	0.093	0.588
XGBoost	458.956	365.439	0.702	0.594	0.666 ± 0.055	0.108	0.580
AdaBoost	445.593	357.625	0.742	0.618	0.677 ± 0.050	0.125	0.604
Bayesian ridge	455.703	361.456	0.700	0.600	0.669 ± 0.058	0.100	0.586
MLP	458.487	365.368	0.702	0.595	0.638 ± 0.090	0.107	0.581
Voting	446.990	351.604	0.779	0.615	0.633 ± 0.026	0.164	0.602
CNN	444.563	350.288	0.624	0.619	0.677 ± 0.041	**0.005**	0.606
DenseNet	455.715	348.658	0.695	0.600	0.677 ± 0.039	0.095	0.586
Stacking	467.695	375.251	0.691	0.579	0.668 ± 0.064	0.112	0.564
AE	458.844	360.194	0.664	0.595	0.666 ± 0.043	0.070	0.580
CatBoost-EY	441.200	–	–	–	–	–	–
ExtRa-EY	449.900	367.000	–	–	–	–	–
RicEns-Net	**437.125**	**335.865**	0.699	0.632	**0.689 ± 0.040**	0.067	**0.619**

in the model. It offers insight into the goodness of fit and serves as an indicator of how effectively the model is likely to predict unseen samples through the explained variance proportion (scikit-learn contributors, 2023).

$$R^2 = 1 - \frac{\sum \left(y_{predicted} - y_{mean}\right)^2}{\sum \left(y_{actual} - y_{mean}\right)^2} \tag{4}$$

Additionally, we employ a modified version of R^2 known as Adjusted R^2, which factors in the impact of an elevated number of predictors contributing to a higher R^2 value. Furthermore, we compute the R^2 score for the input training data to compare it with the testing R^2 score. The disparity between them signals the potential for overfitting. The evaluation of models based on disparities in training and testing R^2 scores aids in the identification of models that may not generalise effectively with unseen data.

$$Adjusted\ R^2 = 1 - \left[\frac{(1 - R^2) \cdot (n - 1)}{(n - k - 1)}\right] \tag{5}$$

where n refers to the number of observations and k is the number of predictors.

4. Results

To comprehensively evaluate the proposed RicEns-Net framework, we conducted a detailed comparative analysis against various baseline models. These include traditional machine learning methods, deep learning models, and ensemble strategies. The evaluation is based on standard regression metrics, namely MAE, RMSE, and the Coefficient of Determination (R^2). Comparative results are reported for both training and independent test datasets to assess the robustness and generalisation performance of RicEns-Net.

We begin by comparing the performance of all state-of-the-art models mentioned above against the proposed RicEns-Net model. Table 4 presents the MAE and RMSE errors in predicting rice yield, along with the goodness of fit measures including train/test and adjusted R^2 values.

Analysing the performance metrics in Table 4, it is evident that the models evaluated for rice crop yield prediction exhibit a range of performance across multiple metrics, including MAE, RMSE, and adjusted R^2. Among these models, the proposed Deep Ensemble model, RicEns-Net, consistently demonstrates more precise predictive capability across all key indicators. Specifically, the RicEns-Net model achieves an MAE of 335.865 kg/Ha and an RMSE of 437.125 kg/Ha, representing improvements of around 3%–5% over the next best-performing models (TabTransformer, SVR and CNN), depending on the metric.

RicEns-Net's standout performance in terms of adjusted R^2, where it achieves a value of 0.619, indicates a better overall fit compared to other models. This is particularly important as adjusted R^2 accounts for the number of predictors in the model, making it a critical measure of the model's generalisation ability. By contrast, while traditional models such as TabTransformer and CNN show relatively strong results, with adjusted R^2 values of 0.614 and 0.606, respectively, they still fall short of RicEns-Net's performance.

Furthermore, SVR and DenseNet, which also rank high in predictive performance, show MAE values of 345.711 and 348.658 kg/Ha, and RMSE values of 448.902 and 455.715 kg/Ha, respectively. However, despite their solid performance, they trail behind RicEns-Net by a margin of approximately 3%–4% in MAE and RMSE.

In the evaluation of predictive performance across models, we performed a 10-fold cross-validation and calculated the average R^2 values for each approach. RicEns-Net demonstrated the highest R^2 score of 0.689 ± 0.040, outperforming other models in terms of

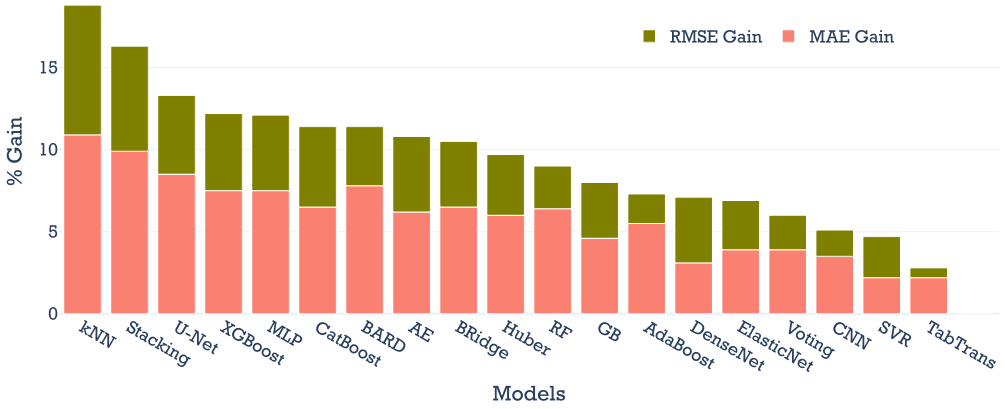

Fig. 6. Percentage performance gain of the RicEns-Net model compared to all other reference models in terms of RMSE and MAE.

predictive accuracy. Among the traditional machine learning models, Bayesian ARD, RF and AdaBoost exhibited strong performance, with R^2 values of 0.678 ± 0.055, 0.678 ± 0.059 and 0.677 ± 0.050, respectively. Similarly, CNN achieved a competitive R^2 score of 0.677 ± 0.041, further validating the efficacy of deep learning architectures in this task. In contrast, models such as Voting Regressor and U-Net yielded lower R^2 scores of 0.633 ± 0.026 and 0.644 ± 0.061, indicating limitations in their ability to handle the complexity of multi-modal data.

Another key observation is the variance in R^2 differences between the train and test sets for each model, highlighting the generalisation strength of each approach. RicEns-Net has the second lowest R^2 differences at 0.067, indicating minimal overfitting compared to other models such as Gradient Boosting (0.209), CatBoost (0.199) and Voting (0.164). This is a significant achievement, as it suggests that RicEns-Net maintains stable performance across both training and unseen test data, further reinforcing its reliability for real-world applications.

While models like AdaBoost, RF, SVR, and voting Regressor also exhibit reasonable performance, with RMSE values ranging from 445 to 449 kg/Ha, and MAE values between 345 and 361 kg/Ha, they are consistently outperformed by RicEns-Net in all metrics. Notably, CatBoost-EY, one of the models from the EY Data Science Challenge, shows a strong RMSE value of 441.200 kg/Ha but lacks data for other metrics, making a full comparison difficult. On the other hand, when comparing the performance of RicEns-Net with the EY Open Science Challenge winning models, the results in Table 4 show that RicEns-Net consistently outperforms these models in terms of MAE and RMSE. Furthermore, models such as TabTransformer, CNN, SVR, DenseNet, and Voting Regressor also display competitive performance compared to the EY winners.

In addition to the quantitative analysis of model performances, Figs. 6, 7, 8, and 9 provide visual representations that further elucidate the effectiveness of RicEns-Net compared to other models.

Fig. 6 presents a stacked bar chart comparing the MAE and RMSE values across all models. This visualisation highlights the performance improvement of RicEns-Net over the alternatives. Notably, TabTransformer, CNN, SVR, and the Voting Regressor models show relatively small performance degradation, with less than a 10% drop in accuracy compared to RicEns-Net. These models are the closest competitors to RicEns-Net in terms of cumulative error metrics. However, models like the kNN, Stacking Regressor, U-Net, and XGBoost exhibit more than a 10% increase in errors, indicating significantly lower predictive accuracy. This graphical representation clearly underscores RicEns-Net's capability in minimising both MAE and RMSE.

Fig. 7 provides a scatter plot that illustrates the relationship between the R^2 Difference (train-test gap) and the adjusted test R^2 values, offering insight into model generalisation and fit. The most effective models, occupying the optimal top-left region of the scatter plot, include RicEns-Net, TabTransformer, CNN, SVR, and ElasticNet. These models not only deliver high adjusted R^2 scores but also maintain small differences between their training and test performances, indicating minimal overfitting and robust generalisation. In contrast, models like the Voting Regressor, CatBoost and AdaBoost, despite finding places in top-performing models in terms of MAE and RMSE, show a significant disparity between training and test R^2 values. This suggests that these models overfit the training data and fail to generalise to unseen data. RicEns-Net's placement in the optimal region confirms its strong generalisation ability and balanced performance across training and test sets.

Fig. 8 presents violin plots visualising the residual distributions (difference between the true and predicted yield values) of the top-performing models. This view allows for a nuanced evaluation of prediction error dispersion, central tendency, and bias. Notably, RicEns-Net and SVR exhibit median residuals close to zero, suggesting well-balanced and unbiased predictions. In contrast, other models, including ElasticNet, TabTransformer, and BARD, tend to have negatively shifted medians, indicating overestimation in yield predictions. While CNN displays the smallest interquartile range (IQR), comparable to RicEns-Net and SVR, it also has noticeably longer tails, implying higher sensitivity to extreme values despite its narrow central error range. An additional pattern emerges across the models from left to right: starting from SVR, residual distributions begin developing a second mode in the upper range. This secondary bulge becomes more visible in ElasticNet, TabTransformer, and BARD, potentially reflecting systematic underprediction for certain observations. Importantly, RicEns-Net and CNN maintain unimodal, symmetric distributions without this pattern. The

Fig. 7. 2D scatterplot of the Adjusted Test R2 and R2 difference values.

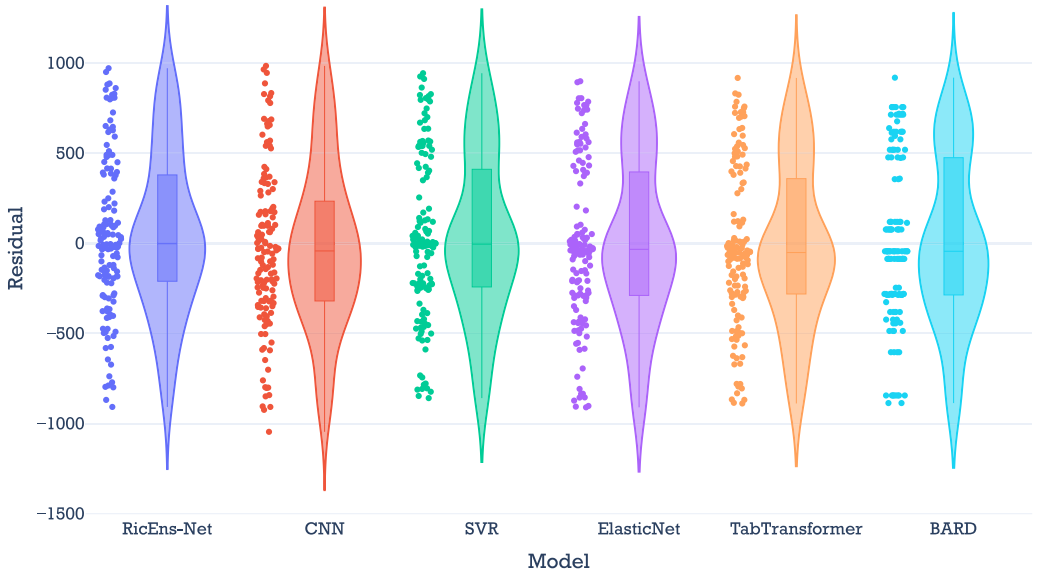

Fig. 8. Residual distributions (true – predicted yield) for the top 6 models.

combination of tight central spread, near-zero median, and unimodal shape confirms RicEns-Net as the most consistent and robust predictor among all tested models.

Finally, Fig. 9 offers a detailed visual analysis of RicEns-Net's predictions for crop yield ($Yield_{rate}$), focusing on spatial error distribution across the test dataset. This figure reveals that, while RicEns-Net maintains consistently low error margins across most of the test regions, there are small areas where the maximum absolute error reaches 10%. These outlier regions are concentrated in areas with higher actual yield rates (darker coloured circles), where RicEns-Net tends to underestimate $Yield_{rate}$. This observation suggests that while RicEns-Net excels in overall yield prediction, further refinement could be beneficial for regions with higher yields. Nevertheless, the majority of the test areas exhibit low prediction errors, further validating the model's reliability in estimating crop yield under diverse conditions.

To sum up the above visual evaluation, the combined results from the figures highlight RicEns-Net's competitive advantage, not only in terms of quantitative metrics but also in its ability to generalise well across different data conditions and accurately predict

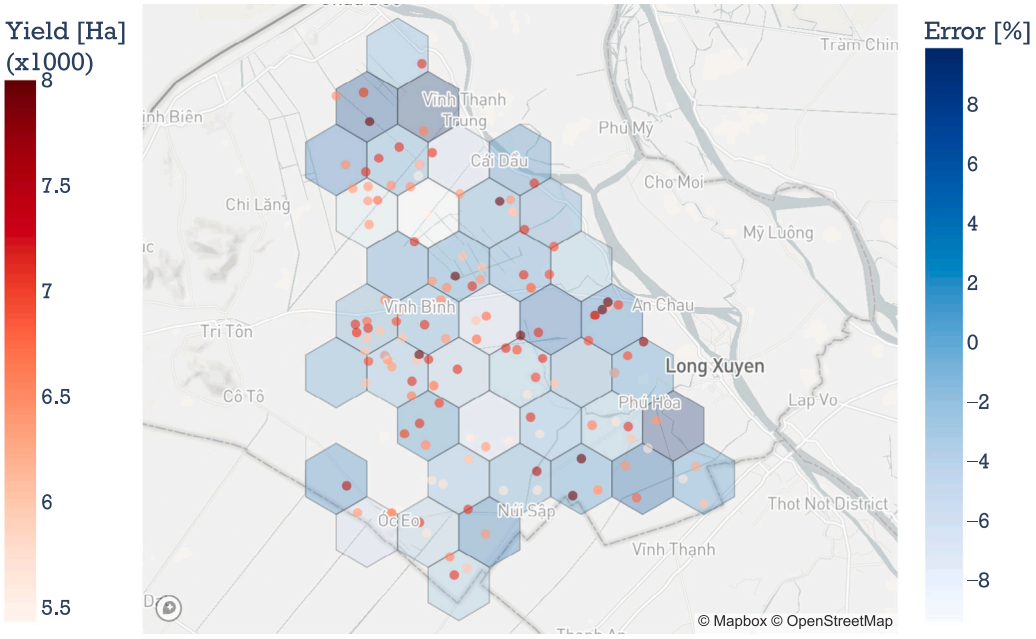

Fig. 9. Map representation of the RicEns-Net's $Yield_{rate}$ prediction error percentages. Coloured hexagons show error percentage averages and dots true Yield values in the test set. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

crop yield with minimal errors across various regions. Both quantitative and qualitative results presented in this section collectively affirm RicEns-Net's effectiveness in yield prediction, highlighting its robustness, precision, and practical applicability in multi-source agricultural modelling.

5. Discussion

The results demonstrate the effectiveness of the proposed RicEns-Net model in achieving high accuracy for crop yield prediction. With a 2% to 10% improvement in both MAE and RMSE compared to other state-of-the-art models, RicEns-Net clearly shows its capability, particularly in handling multi-modal data from radar, optical, and meteorological sources. This aligns with previous findings where deep ensemble models have demonstrated enhanced predictive capabilities through the combination of diverse feature sets (Abdali et al., 2024; Shahhosseini et al., 2020; Keerthana et al., 2021). The significant improvement in adjusted R^2 further emphasises the robustness of RicEns-Net, which surpasses even advanced models such as Transformers, CNN and SVR.

A closer inspection of the performance of traditional ensemble methods like the Voting Regressor, or boosting methods of Ada- and CatBoost reveals their strong performance in MAE and RMSE metrics. However, these models exhibit overfitting tendencies, as indicated by the large R^2 disparities between training and test phases. This overfitting issue highlights a major advantage of RicEns-Net, which maintains a balance between training and test performance, avoiding overfitting through its deep ensemble design.

The performance of the RicEns-Net model, achieving a mean absolute error (MAE) of 336 kg/Ha, demonstrates its robust predictive capability in the context of crop yield forecasting for the Mekong Delta region in Vietnam. This region, often referred to as Vietnam's "rice bowl", contributes more than half of the nation's rice output and is critical to ensuring both domestic food security and Vietnam's position as a leading global rice exporter (Food and Agriculture Organization (FAO), 2022; Maitah et al., 2020). Given the region's susceptibility to climate variability, such as temperature fluctuations and erratic rainfall patterns, achieving high predictive accuracy in yield estimates is essential for implementing timely interventions and optimising resource allocation. The accuracy of 336 kg/Ha represents a meaningful advancement, given that the lowest average yield in the region is approximately 6000 kg/Ha, as reported by Clauss et al. (2018). This error equates to around 5%–6% of the lowest yield, reaching a maximum of 10% in some cases, as illustrated in Fig. 9. These findings underscore the value of RicEns-Net in enhancing crop yield predictions and its potential for broader applications in precision agriculture.

Furthermore, when comparing the results to the EY Open Science Challenge 2023 winners, the competitive performance of RicEns-Net is evident. This further validates the efficacy of the proposed deep ensemble approach for predictive tasks relying on multi-modal data. While models like TabTransformer, CNN and SVR also achieve competitive results, the effectiveness of neural network-based architectures for large-scale agricultural datasets is underscored. However, the underestimation of yield rates in high-yield regions, as noted in Fig. 9, suggests that future work could explore refining the model for these specific outlier cases.

6. Conclusions

This study introduced **RicEns-Net**, a novel deep ensemble model for rice yield prediction, integrating multi-modal data sources including synthetic aperture radar (SAR), optical remote sensing, and meteorological measurements. Through comprehensive feature engineering and ensemble learning strategies, RicEns-Net consistently outperformed traditional machine learning models and recent state-of-the-art methods across various performance metrics (MAE, RMSE, R^2). Our experiments, conducted on real-world data from the Mekong Delta region, demonstrated that leveraging diverse data modalities and ensemble approaches can substantially enhance the accuracy and robustness of crop yield forecasts. The findings contribute to the growing body of research advocating for integrated remote sensing and machine learning approaches in precision agriculture.

6.1. Limitations

Despite the promising results, several limitations of this study should be acknowledged. First, biases in the data could arise from the quality and availability of remote sensing and meteorological inputs. For example, cloud cover in multispectral imagery may introduce spatial and temporal gaps, potentially affecting model training and prediction accuracy in certain regions. Second, although RicEns-Net demonstrated robust performance across the Mekong Delta's environmental conditions, its generalisability to other regions with different climate patterns, crop types, or farming practices remains to be fully validated. Third, the model assumes that the selected features sufficiently capture the variability in yield outcomes, yet site-specific factors such as soil health, irrigation practices, or pest pressures may not be fully represented. These assumptions highlight areas for future refinement to ensure broader applicability and accuracy across diverse agricultural contexts.

6.2. Practical implications

The practical implications of RicEns-Net for precision agriculture are significant. By accurately predicting crop yields based on multi-modal remote sensing and meteorological data, farmers and agricultural planners can make more informed decisions regarding resource allocation. For instance, optimised irrigation schedules and targeted fertiliser applications based on predicted yield potential can enhance both efficiency and sustainability. The use of radar and optical data enables real-time monitoring of crop and soil health, empowering farmers to take proactive measures against yield losses due to drought, disease, or pest outbreaks. Furthermore, RicEns-Net's integration of meteorological information provides valuable risk management insights by anticipating weather-driven yield variability. As remote sensing technologies become increasingly accessible and affordable, models like RicEns-Net offer the potential for seamless integration into farm management systems, providing intuitive, data-driven decision support tools that can contribute to increased productivity and resilience in agricultural operations.

6.3. Future work

Future research could extend this study in several important directions. Enhancing the temporal resolution of input data remains a priority, particularly under conditions where satellite acquisitions are hindered by weather. Integrating UAV-based hyperspectral and LiDAR data, with their higher flexibility and spatial detail, could significantly refine yield predictions and enable finer-scale management recommendations. Moreover, emerging machine learning architectures, such as transformer models and self-supervised learning techniques, offer promising avenues for improving predictive performance and learning from sparse or incomplete data. Another area of improvement lies in incorporating additional ground-truth data, such as detailed soil health measurements or farm management records, to better capture localised influences on yield variability. Finally, expanding the application of RicEns-Net to different crops and diverse agro-ecological zones would allow for a robust assessment of its generalisability and practical value at broader scales.

CRediT authorship contribution statement

Akshay Dagadu Yewle: Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Laman Mirzayeva:** Writing – review & editing, Validation, Investigation, Formal analysis. **Oktay Karakuş:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Methodology, Funding acquisition, Formal analysis, Conceptualization.

Ethics in publishing statement

This research presents an accurate account of the work performed, all data presented are accurate and methodologies are detailed enough to permit others to replicate the work.

This manuscript represents entirely original works and or, if work and/or words of others have been used, that this has been appropriately cited or quoted, and permission has been obtained where necessary.

This material has not been published in whole or in part elsewhere.

The manuscript is not currently being considered for publication in another journal.

All authors have been personally and actively involved in substantive work leading to the manuscript and will hold themselves jointly and individually responsible for its content.

Declaration of Generative AI and AI-assisted technologies in the writing process

That generative AI and AI-assisted technologies have not been utilised in the writing process or, if used, disclosed in the manuscript the use of AI and AI-assisted technologies, and a statement will appear in the published work.

That generative AI and AI-assisted technologies have not been used to create or alter images unless specifically used as part of the research design, where such use must be described in a reproducible manner in the methods section.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix. Vegetation, soil and plant biochemical indices

This appendix section presents important vegetation, soil and plant biochemical indices in Table A.5.

Data availability

The authors do not have permission to share data.

Table A.5
Vegetation, Soil, Water and Plant biochemical indices.

Index	Reference
Normalised Difference Vegetation Index (**NDVI**)	$\frac{R_{NIR} - R_{RED}}{R_{NIR} + R_{RED}}$
Transformed Vegetation Index (**TVI**)	$\sqrt{NDVI + 0.5}$
Simple Ratio (**SR**)	$\frac{R_{NIR}}{R_{RED}}$
Enhanced Vegetation Index (**EVI**)	$2.5 \times \frac{R_{NIR} - R_{RED}}{R_{NIR} + 6R_{RED} - 7.5R + 1}$
EVI - 2-Bands (**EVI2**)	$2.5 \times \frac{R_{NIR} - R_{RED}}{R_{NIR} + 2.4R_{RED} + 1}$
Soil adjusted vegetation index (**SAVI**)	$1.5 \times \frac{R_{NIR} - R_{RED}}{R_{NIR} + R_{RED} + 0.5}$
Rice Growth Vegetation Index (**RGVI**)	$1 - \left(\frac{R + R_{RED}}{R_{NIR} + R_{SWIR1} + R_{SWIR2}} \right)$
Difference Vegetation Index (**DVI**)	$R_{NIR} - R_{RED}$
Modified Simple Ratio (**MSR**)	$\frac{SR - 1}{\sqrt{SR + 1}}$
Near Infra-Red Reflectance Of Vegetation (**NIRv**)	$NDVI \times R_{NIR}$
Kernelized NDVI (**kNDVI**)	$\tanh(NDVI^2)$
NDVI-Red Edge (NNDVIre)	$\frac{R_{NIR} - R_{RedEdge1}}{R_{NIR} + R_{RedEdge1}}$
Normalised Difference Red Edge 1 (**NDRE1**)	$\frac{R_{RedEdge2} - R_{RedEdge1}}{R_{RedEdge2} + R_{RedEdge1}}$
Normalised Difference Red Edge 2 (**NDRE2**)	$\frac{R_{RedEdge3} - R_{RedEdge1}}{R_{RedEdge3} + R_{RedEdge1}}$
Normalised Difference Water Index (**NDWI**)	$\frac{R_{GREEN} - R_{NIR}}{R_{GREEN} + R_{NIR}}$
Bare Soil Index (**BSI**)	$\frac{(R_{RED} + R_{SWIR1}) - (R_{NIR} + R)}{(R_{RED} + R_{SWIR1}) + (R_{NIR} + R)}$
Land Surface Water Index (1.6 μm) (**LSWI16**)	$\frac{R_{NIR} - R_{SWIR1}}{R_{NIR} + R_{SWIR1}}$
Land Surface Water Index (2.2 μm) (**LSWI22**)	$\frac{R_{NIR} - R_{SWIR2}}{R_{NIR} + R_{SWIR2}}$
Chlorophyll Carotenoid Index (**CCI**)	$\frac{R_{GREEN} - R_{RED}}{R_{GREEN} + R_{RED}}$
Green Chromatic Coordinate (**GCC**)	$\frac{R_{GREEN}}{R_{RED} + R_{GREEN} + R}$

References

Abdali, E., Valadan Zoej, M.J., Taheri Dehkordi, A., Ghaderpour, E., 2024. A parallel-cascaded ensemble of machine learning models for crop type classification in google earth engine using multi-temporal sentinel-1/2 and landsat-8/9 remote sensing data. Remote. Sens. 16 (1), 127.

Akhtarmanesh, A., Abbasi-Moghadam, D., Sharifi, A., Yadkouri, M.H., Tariq, A., Lu, L., 2023. Road extraction from satellite images using attention-assisted UNet. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 17, 1126–1136.

Akiyama, K., Alberdi, A., Alef, W., Asada, K., Azulay, R., Baczko, A.-K., Ball, D., Baloković, M., Barrett, J., Bintley, D., et al., 2019. First M87 event horizon telescope results. IV. Imaging the central supermassive black hole. Astrophys. J. Lett. 875 (1), L4.

Arshad, S., Kazmi, J.H., Javed, M.G., Mohammed, S., 2023. Applicability of machine learning techniques in predicting wheat yield based on remote sensing and climate data in Pakistan, South Asia. Eur. J. Agron. 147, 126837.

Asadollah, S.B.H.S., Jodar-Abellan, A., Pardo, M.Á., 2024. Optimizing machine learning for agricultural productivity: A novel approach with rscv and remote sensing data over Europe. Agricult. Sys. 218, 103955.

Baier, W., Robertson, G.W., 1968. The performance of soil moisture estimates as compared with the direct use of climatological data for estimating crop yields. Agric. Meteorol. 5 (1), 17–31.

Bala, S., Islam, A., 2009. Correlation between potato yield and MODIS-derived vegetation indices. Int. J. Remote Sens. 30 (10), 2491–2507.

Baret, F., Guyot, G., 1991. Potentials and limits of vegetation indices for LAI and APAR assessment. Remote Sens. Environ. 35 (2–3), 161–173.

Biermann, L., Clewley, D., Martinez-Vicente, V., Topouzelis, K., 2020. Finding plastic patches in coastal waters using optical satellite data. Sci. Rep. 10 (1), 5364.

Booth, H., Ma, W., Karakuş, O., 2023. High-precision density mapping of marine debris and floating plastics via satellite imagery. Sci. Rep. 13 (1), 6822.

Boppudi, S., et al., 2024. Deep ensemble model with hybrid intelligence technique for crop yield prediction. Multimedia Tools Appl. 1–21.

Bouman, B., 1995. Crop modelling and remote sensing for yield prediction. Neth. J. Agric. Sci. 43 (2), 143–161.

Chen, R., Zhang, C., Xu, B., Zhu, Y., Zhao, F., Han, S., Yang, G., Yang, H., 2022. Predicting individual apple tree yield using UAV multi-source remote sensing data and ensemble learning. Comput. Electron. Agric. 201, 107275. http://dx.doi.org/10.1016/j.compag.2022.107275, URL https://www.sciencedirect.com/science/article/pii/S0168169922005877.

Clauss, K., Ottinger, M., Leinenkugel, P., Kuenzer, C., 2018. Estimating rice production in the Mekong Delta, Vietnam, utilizing time series of Sentinel-1 SAR data. Int. J. Appl. Earth Obs. Geoinf. 73, 574–585. http://dx.doi.org/10.1016/j.jag.2018.07.022.

Dhaliwal, D.S., Williams, M.M., 2024. Sweet corn yield prediction using machine learning models and field-level data. Precis. Agric. 25 (1), 51–64.

e-Extension, 2024. Major insect pests at each growth stage of the rice plant. URL https://irri-e-extension.com/lesson/major-insect-pests-at-each-growth-stage-of-the-rice-plant/. (Accessed 20 January 2024).

Esmaeili, M., Abbasi-Moghadam, D., Sharifi, A., Tariq, A., Li, Q., 2023. ResMorCNN model: hyperspectral images classification using residual-injection morphological features and 3DCNN layers. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 17, 219–243.

EY, 2023. Open science data challenge. URL https://challenge.ey.com/. (Accessed 20 August 2023).

Farmonov, N., Esmaeili, M., Abbasi-Moghadam, D., Sharifi, A., Amankulova, K., Mucsi, L., 2024. HypsLiDNet: 3-D–2-D CNN model and spatial–spectral morphological attention for crop classification with DESIS and LiDAR data. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 17, 11969–11996. http://dx.doi.org/10.1109/JSTARS.2024.3418854.

Food and Agriculture Organization (FAO), 2022. Rice market monitor. URL https://www.fao.org/3/ca0412en/CA0412EN.pdf. (Accessed January 18, 2024).

Gadupudi, A., Rani, R.Y., Jayaram, B., Sharma, N., Deshmukh, J.K., Reddy, P.C.S., 2024. An adaptive deep learning model for crop yield prediction. In: 2024 2nd International Conference on Computer, Communication and Control (IC4). IEEE, pp. 1–5.

Gavahi, K., Abbaszadeh, P., Moradkhani, H., 2021. DeepYield: A combined convolutional neural network with long short-term memory for crop yield forecasting. Expert Syst. Appl. 184, 115511.

Gopi, P., Karthikeyan, M., 2024. Red fox optimization with ensemble recurrent neural network for crop recommendation and yield prediction model. Multimedia Tools Appl. 83 (5), 13159–13179.

Hashemi, M.G., Tan, P.-N., Jalilvand, E., Wilke, B., Alemohammad, H., Das, N.N., 2024. Yield estimation from SAR data using patch-based deep learning and machine learning techniques. Comput. Electron. Agric. 226, 109340.

Hussain, K., Mehmood, K., Yujun, S., Badshah, T., Anees, S.A., Shahzad, F., Nooruddin, Ali, J., Bilal, M., 2024. Analysing LULC transformations using remote sensing data: insights from a multilayer perceptron neural network approach. Ann. GIS 1–28.

Islam, M.D., Di, L., Qamer, F.M., Shrestha, S., Guo, L., Lin, L., Mayer, T.J., Phalke, A.R., 2023. Rapid rice yield estimation using integrated remote sensing and meteorological data and machine learning. Remote. Sens. 15 (9), 2374.

Johnson, M.D., Hsieh, W.W., Cannon, A.J., Davidson, A., Bédard, F., 2016. Crop yield forecasting on the Canadian Prairies by remotely sensed vegetation indices and machine learning methods. Agricult. Forest. Meterol. 218, 74–84.

Joshi, A., Pradhan, B., Gite, S., Chakraborty, S., 2023. Remote-sensing data and deep-learning techniques in crop mapping and yield prediction: A systematic review. Remote. Sens. 15 (8), 2014.

Keerthana, M., Meghana, K., Pravallika, S., Kavitha, M., 2021. An ensemble algorithm for crop yield prediction. In: 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks. ICICV, IEEE, pp. 963–970.

Khan, H.R., Gillani, Z., Jamal, M.H., Athar, A., Chaudhry, M.T., Chao, H., He, Y., Chen, M., 2023. Early identification of crop type for smallholder farming systems using deep learning on time-series Sentinel-2 imagery. Sensors 23 (4), 1779.

Killeen, P., Kiringa, I., Yeap, T., Branco, P., 2024. Corn grain yield prediction using UAV-based high spatiotemporal resolution imagery, machine learning, and spatial cross-validation. Remote. Sens. 16 (4), 683.

Li, R., Li, C., Xu, X., Wang, J., Yang, X., Huang, W., Pan, Y., et al., 2009. Winter wheat yield estimation based on support vector machine regression and multi-temporal remote sensing data. Trans. Chin. Soc. Agric. Eng. 25 (7), 114–117.

Li, A., Liang, S., Wang, A., Qin, J., 2007. Estimating crop yield from multi-temporal satellite data using multivariate regression and neural network techniques. Photogramm. Eng. Remote Sens. 73 (10), 1149–1157.

Li, X., Yuan, W., Dong, W., 2021. A machine learning method for predicting vegetation indices in China. Remote. Sens. 13 (6), 1147.

Li, Y., Zeng, H., Zhang, M., Wu, B., Qin, X., 2024. Global de-trending significantly improves the accuracy of XGBoost-based county-level maize and soybean yield prediction in the Midwestern United States. GIScience Remote. Sens. 61 (1), 2349341.

Lu, J., Fu, H., Tang, X., Liu, Z., Huang, J., Zou, W., Chen, H., Sun, Y., Ning, X., Li, J., 2024. GOA-optimized deep learning for soybean yield estimation using multi-source remote sensing data. Sci. Rep. 14 (1), 7097.

Ma, W., Karakuş, O., Rosin, P.L., 2022. AMM-FuseNet: Attention-based multi-modal image fusion network for land cover mapping. Remote. Sens. 14 (18), 4458.

Ma, W., Karakuş, O., Rosin, P.L., 2024. Knowledge distillation for road detection based on cross-model semi-supervised learning. In: IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 8173–8178.

Mahdipour, H., Sharifi, A., Sookhak, M., Medrano, C.R., 2024. Ultrafusion: Optimal fuzzy fusion in land-cover segmentation using multiple panchromatic satellite images. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 17, 5721–5733. http://dx.doi.org/10.1109/JSTARS.2024.3360648.

Maimaitijiang, M., Sagan, V., Sidike, P., Hartling, S., Esposito, F., Fritschi, F.B., 2020. Soybean yield prediction from UAV using multimodal data fusion and deep learning. Remote Sens. Environ. 237, 111599.

Maitah, K., Smutka, L., Sahatqija, J., Maitah, M., Phuong Anh, N., 2020. Rice as a determinant of Vietnamese economic sustainability. Sustainability 12 (12), http://dx.doi.org/10.3390/su12125123.

Manafifard, M., 2024. A new hyperparameter to random forest: application of remote sensing in yield prediction. Earth Sci. Inform. 17 (1), 63–73.

Marzvan, S., Moravej, K., Felegari, S., Sharifi, A., Askari, M.S., 2021. Risk assessment of alien azolla filiculoides lam in anzali lagoon using remote sensing imagery. J. Indian Soc. Remote. Sens. 49, 1801–1809.

Mirhoseini Nejad, S.M., Abbasi-Moghadam, D., Sharifi, A., 2024. ConvLSTM–Vit: A deep neural network for crop yield prediction using earth observations and remotely sensed data. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 17, 17489–17502. http://dx.doi.org/10.1109/JSTARS.2024.3464411.

Mosleh, M.K., Hassan, Q.K., Chowdhury, E.H., 2015. Application of remote sensors in mapping rice area and forecasting its production: A review. Sensors 15 (1), 769–791.

Osibo, B.K., Ma, T., Wahab, M.M.A., Jia, L., Ye, W., Bediako-Kyeremeh, B., Mamelona, L., Darbinian, K., 2024. Integrating remote sensing data and efficiently weighted ensemble method for simplified crop yield prediction. J. Appl. Remote. Sens. 18 (4), 044520.

Pantazi, X.E., Moshou, D., Alexandridis, T., Whetton, R.L., Mouazen, A.M., 2016. Wheat yield prediction using machine learning and advanced sensing techniques. Comput. Electron. Agric. 121, 57–65.

Perich, G., Turkoglu, M.O., Graf, L.V., Wegner, J.D., Aasen, H., Walter, A., Liebisch, F., 2023. Pixel-based yield mapping and prediction from Sentinel-2 using spectral indices and neural networks. Field Crop. Res. 292, 108824.

Ramos, A.P.M., Osco, L.P., Furuya, D.E.G., Gonçalves, W.N., Santana, D.C., Teodoro, L.P.R., da Silva Junior, C.A., Capristo-Silva, G.F., Li, J., Baio, F.H.R., et al., 2020. A random forest ranking approach to predict yield in maize with UAV-based vegetation spectral indices. Comput. Electron. Agric. 178, 105791.

Rao, M.V., Sreeraman, Y., Mantena, S.V., Gundu, V., Roja, D., Vatambeti, R., 2024. Brinjal Crop yield prediction using Shuffled shepherd optimization algorithm based ACNN-OBDLSTM model in Smart Agriculture. J. Integr. Sci. Technol. 12 (1), 710.

Rowley, A., Karakuş, O., 2023. Predicting air quality via multimodal AI and satellite imagery. Remote Sens. Environ. 293, 113609.

Roy, S.K., Deria, A., Hong, D., Rasti, B., Plaza, A., Chanussot, J., 2023. Multimodal fusion transformer for remote sensing image classification. IEEE Trans. Geosci. Remote Sens..

Safari, M.M., Sharifi, A., Mahmood, J., Abbasi-Moghadam, D., 2024. Mesoscale eddy detection and classification from sea surface temperature maps with deep neural networks. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens..

scikit-learn contributors, 2023. Scikit-learn - R2 score documentation. URL https://scikit-learn.org/stable/modules/generated/sklearn.metrics.r2_score.html. (Accessed 3 October 2023).

Shafi, U., Mumtaz, R., Anwar, Z., Ajmal, M.M., Khan, M.A., Mahmood, Z., Qamar, M., Jhanzab, H.M., 2023. Tackling food insecurity using remote sensing and machine learning based crop yield prediction. IEEE Access.

Shahhosseini, M., Hu, G., Archontoulis, S.V., 2020. Forecasting corn yield with machine learning ensembles. Front. Plant Sci. 11, 527890.

Shahhosseini, M., Hu, G., Khaki, S., Archontoulis, S.V., 2021. Corn yield prediction with ensemble CNN-DNN. Front. Plant Sci. 12, 709008.

Sharma, S., Jain, A., Sharma, S., Whig, P., 2025. Enhancing crop yield prediction through machine learning regression analysis. Int. J. Sustain. Agric. Manag. Informatics 11 (1), 29–47.

Son, N.-T., Chen, C.-F., Cheng, Y.-S., Toscano, P., Chen, C.-R., Chen, S.-L., Tseng, K.-H., Syu, C.-H., Guo, H.-Y., Zhang, Y.-T., 2022. Field-scale rice yield prediction from sentinel-2 monthly image composites using machine learning algorithms. Ecol. Inform. 69, 101618.

Stojanova, D., Panov, P., Gjorgjioski, V., Kobler, A., Džeroski, S., 2010. Estimating vegetation height and canopy cover from remotely sensed data with machine learning. Ecol. Inform. 5 (4), 256–266.

Tao, W., Zeng, S., Su, L., Sun, Y., Shao, F., Wang, Q., 2023. Yield prediction and water-nitrogen management of Chinese jujube based on machine learning. Irrig. Drain. 72 (2), 439–450. http://dx.doi.org/10.1002/ird.2786.

Trentin, C., Ampatzidis, Y., Lacerda, C., Shiratsuchi, L., 2024. Tree crop yield estimation and prediction using remote sensing and machine learning: A systematic review. Smart Agric. Technol. 100556.

Umamaheswari, K., Madhumathi, R., 2024. Predicting crop yield based on stacking ensemble model in machine learning. In: 2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC). IEEE, pp. 1831–1836.

United Nations, 2023a. Goal 2: Zero hunger. n.d. URL https://sdgs.un.org/goals/goal2. (Accessed 20 August 2023).

United Nations, 2023b. United nations sustainable development goals. URL https://sdgs.un.org/goals. (Accessed 20 August 2023).

Uno, Y., Prasher, S., Lacroix, R., Goel, P., Karimi, Y., Viau, A., Patel, R., 2005. Artificial neural networks to predict corn yield from Compact Airborne Spectrographic Imager data. Comput. Electron. Agric. 47 (2), 149–161.

Wu, R., Fan, Y., Zhang, L., Yuan, D., Gao, G., 2024. Wheat yield estimation study using hyperspectral vegetation indices. Appl. Sci. 14 (10), 4245.

Zare, H., Weber, T.K., Ingwersen, J., Nowak, W., Gayler, S., Streck, T., 2024. Within-season crop yield prediction by a multi-model ensemble with integrated data assimilation. Field Crop. Res. 308, 109293.

Zhang, P., Lu, B., Shang, J., Wang, X., Hou, Z., Jin, S., Yang, Y., Zang, H., Ge, J., Zeng, Z., 2024. Ensemble learning for oat yield prediction using multi-growth stage UAV images. Remote. Sens. 16 (23), 4575.

Zhang, L., Zhang, Z., Luo, Y., Cao, J., Xie, R., Li, S., 2021. Integrating satellite-derived climatic and vegetation indices to predict smallholder maize yield using deep learning. Agricult. Forest. Meterol. 311, 108666.

Zhou, S., Xu, L., Chen, N., 2023. Rice yield prediction in hubei province based on deep learning and the effect of spatial heterogeneity. Remote. Sens. 15 (5), 1361.