STAF-Net: An Innovative Framework for Wheat Yield Prediction

Mohamed El Ayyadi[1], Khadija Meghraoui[1], Maryam El Hamdani[2], Saloua Bensiali[2,3] Assia Lozzi[2,4], Imane Sebari [1,5]

[1] Research Unit of Geospatial Technologies for a Smart Decision, IAV Hassan II, Rabat, 10101, Morocco
[2] Research Unit of Plant Production Sciences and Agroecology, IAV Hassan II, Rabat, 10101, Morocco
[3] Department of Applied Statistics and Computer Science, IAV Hassan II, Rabat, 10101, Morocco
[4] Department of Crop Production, Protection and Biotechnology, IAV Hassan II, Rabat, 10101, Morocco
[5] Department of Photogrammetry and Cartography, School of Geomatics and Surveying Engineering, IAV Hassan II, Rabat 10101, Morocco

KEY WORDS: wheat yield prediction; Sentinel-2; STAFNet; GAN; machine learning; deep learning.

ABSTRACT:

Accurate crop yield forecasting is critical for optimizing agricultural resource management and ensuring food security. This study introduces STAFNet (Spatial-Temporal Attention Fusion Network), an innovative deep learning framework designed to integrate multispectral Sentinel-2 imagery and climatic variables for wheat yield prediction under limited data conditions. Classical machine learning models (Random Forest, XGBoost, Support Vector Machine) and a CNN-LSTM architecture were evaluated for comparison. Additionally, a Generative Adversarial Network (GAN) was employed to generate realistic synthetic multispectral images, addressing dataset scarcity and enhancing model generalization. Experiments were conducted in Sidi Yahya Zaer, Morocco, using simulated yield data derived from NDVI-based statistical modeling for the 2020–2024 period. Results show that XGBoost achieved strong baseline performance ($R^2 = 0.919$), while STAFNet exhibited superior temporal stability and accuracy. Incorporating GAN-based augmentation further improved STAFNet's performance, reaching $R^2 = 0.935$ and significantly reducing RMSE and MAE. Multi-horizon testing confirmed robust early-season predictive capability from January onwards. These findings highlight the combined benefits of attention-based architectures and synthetic data generation for in-season yield forecasting, offering a scalable, cost-effective solution adaptable to various crops and regions.

1. Introduction

The agricultural sector faces numerous challenges in achieving accurate crop yield estimation, which is essential for effective resource management and addressing food shortages in a rapidly growing global population (Jabed and Azmi Murad, 2024). Reliable yield forecasting also supports decision-making related to agricultural trade (Cunha and Silva, 2020).

Recent advancements in remote sensing have greatly improved the ability to monitor and predict crop yields. These technologies provide valuable data on crop health, soil, and environmental conditions, enabling more accurate and timely estimates of agricultural productivity (Franz et al., 2020). Among them, the Sentinel satellite family, developed by the European Space Agency (ESA) under the Copernicus program, plays a pivotal role. Launched in 2015, Sentinel-2 (S2) overcomes previous limitations with its high spatial (10–60 m), spectral (13 bands), and temporal (five-day revisit) resolutions. Equipped with a multispectral instrument (MSI), S2 captures data across visible, near-infrared, and shortwave infrared bands, allowing detailed monitoring of crop growth throughout the season (Zhao et al., 2020). Sentinel-2 is particularly valuable for subsistence agriculture, supporting precise crop health assessments and yield predictions (Aslan et al., 2024).

Many studies have explored different approaches for crop yield prediction (Aslan et al., 2024). Traditional statistical models rely on historical yield and environmental data (Burdett and Wellen, 2022), while advanced machine learning algorithms such as Random Forest (RF), Support Vector Machines (SVM), and XGBoost capture complex nonlinear relationships (Bouras et al., 2021). Deep learning methods can automatically extract features and model long-term dependencies in time-series data, improving accuracy and robustness (Krithika and Sangeetha, 2024). Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks have been increasingly applied to crop yield prediction for their strengths in handling spatial and temporal information (Gavahi et al., 2021). Combined CNN-LSTM architectures show strong potential by leveraging both aspects effectively (Krithika and Sangeetha, 2024). More recently, Vision Transformers (ViT) have emerged as a promising alternative, using self-attention mechanisms to capture global relationships between image patches, potentially enhancing spatial modeling performance (Dosovitskiy et al., 2021; Wang et al., 2025; Rodrigo et al., 2024).

Building upon the Vision Transformer framework, the Pyramid Vision Transformer (PVT) was developed specifically for computer vision tasks. Unlike the standard ViT, which produces fixed-resolution global representations, PVT incorporates a hierarchical pyramid structure similar to that of CNNs his design progressively processes the image through multiple stages, reducing spatial dimensions while increasing channel depth, allowing for more efficient computation. Additionally, PVT employs a spatial-reduction attention mechanism that lowers computational costs and memory usage, making it well-suited for handling high-resolution images (Wang et al., 2021).

Data augmentation is used to compensate for limited training data in deep learning, especially in computer vision, by applying geometric and color transformations that expand datasets and improve model generalization (Shorten and Khoshgoftaar, 2019). Generative Adversarial Networks (GANs), introduced by Goodfellow et al. (2014), generate realistic synthetic images by learning real data distributions (Goodfellow et al., 2020). In yield prediction, GANs help mitigate data scarcity and class imbalance, particularly in agriculture where data collection is expensive and time-consuming (Zhang et al., 2022). They consist of a generator creating fake samples and a discriminator distinguishing them from real ones until both reach equilibrium (Pan et al., 2019).

The main contributions of this work are summarized as follows:
• We propose STAF-Net, a novel spatiotemporal-attention-based framework that integrates multispectral and climatic data for wheat yield prediction.
• We introduce a GAN-based data augmentation strategy to overcome data scarcity in agricultural contexts.
• We evaluate the proposed model on multisource Sentinel-2 and climate data for Morocco (2020–2024), showing its robustness compared to traditional ML and DL baselines.

2. Methodology

The methodological framework adopted in this study is illustrated in Figure 1.

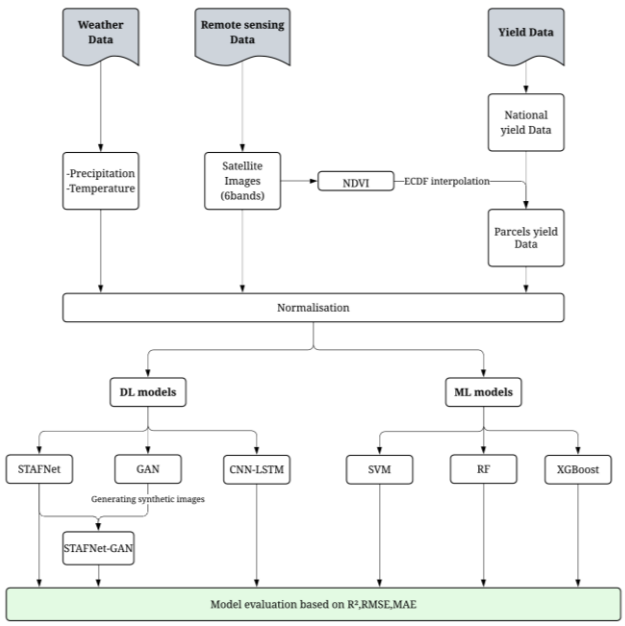

Figure 1. The study methodological workflow.

It integrates multi-source datasets, including climatic variables, remote sensing imagery, and yield statistics, into a unified predictive modeling pipeline. Weather data (precipitation and temperature) and Sentinel-2 multispectral images (six selected spectral bands) were processed alongside national yield records. The satellite imagery was used both to compute the Normalized Difference Vegetation Index (NDVI) for yield data simulation via ECDF-based interpolation, and directly as multi-band spectral inputs for the prediction models. All datasets underwent normalization to ensure comparability and stability during model training. Two modeling families were evaluated: (i) deep learning models, including STAFNet, CNN-LSTM, and a hybrid STAFNet-GAN architecture incorporating synthetic image generation; and (ii) machine learning models, including Support Vector Machines (SVM), Random Forest (RF), and XGBoost. Model performance was assessed using three key metrics coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE) to enable a robust comparative analysis across approaches.

2.1 Study Area

The study was conducted in the commune of Sidi Yahya Zaer as illustrated in Figure 2, located in the Rabat-Salé-Kénitra region,Morocco, is characterized by significant natural heterogeneity. The area exhibits a diverse topography shaped by a Mediterranean climate, influenced by its proximity to the Atlantic Ocean. These environmental conditions create a favorable setting for agriculture, particularly cereal cultivation, which plays a vital role in the region's agricultural landscape.Due to its topographical and climatic diversity, Sidi Yahya Zaer serves as a relevant study area for crop yield prediction. It offers the opportunity to analyze crop performance under varied conditions, making it representative of many agroecological contexts across Morocco. The study area comprises 30 agricultural parcels,which depicts their geographical distribution within the commune.

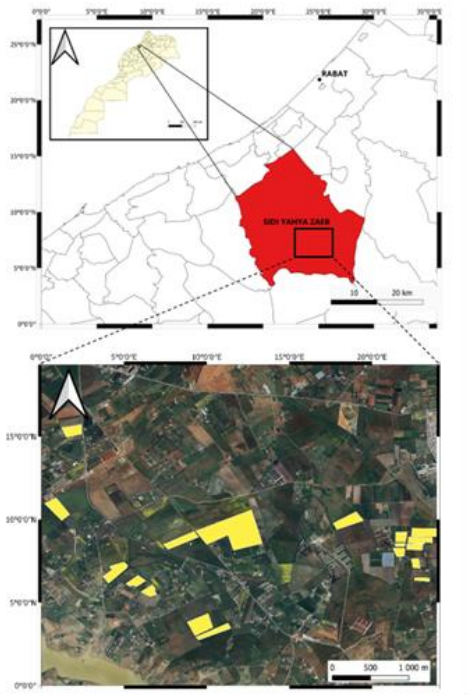

Figure 2. Study area

2.2 Data

2.2.1 Remote sensing Data

In this study, remote sensing data constituted the core input for monitoring crop growth and predicting wheat yields. Sentinel-2 imagery was selected for its high spatial and temporal resolution, as well as its free accessibility. The Sentinel-2 mission, developed by the European Space Agency (ESA), provides multispectral optical data with 13 spectral bands at resolutions ranging from 10 to 60 meters. Specifically, the surface reflectance product (S2_SR) was utilized, which offers atmospherically corrected reflectance values for the period spanning 2020 to 2024 and were retrieved using the Google Earth Engine (GEE) platform. This platform provides fast and efficient access to high-resolution time-series imagery (Gorelick et al., 2017).
Image extraction was specifically conducted for the months corresponding to the wheat growth cycle, from January to May. In total, 750 images were extracted (each containing 6 bands), covering 30 parcels over 5 years, with one image per month.

2.2.2 Weather Data

To enhance the accuracy of wheat yield predictions, meteorological variables, specifically precipitation and temperature, were incorporated into the model for the critical months of November and December (year N-1) and January (year N). The selection of these variables is supported by prior research, notably (Bouras et al., 2021), which established a significant correlation between climatic factors and cereal yields

in Morocco. According to Bouras et al., abundant rainfall during November and December promotes rapid plant emergence and expands cultivated areas, leading to improved overall production. Similarly, a positive correlation between temperature and yield was observed during December and January, aligning with the early growth stages of wheat. Climatic data were sourced from NASA POWER, with precipitation data derived from the CHIRPS (Climate Hazards Group InfraRed Precipitation with Station data) dataset, known for its high resolution and reliability in capturing rainfall patterns. The selected variables include total precipitation (mm) and average temperature (°C), systematically organized to cover November and December of the previous year (N-1) and January of the study year (N), aligning with the wheat growth cycle's critical phenological stages.

2.2.3 Wheat Yield Data

To model wheat yield predictions, a dataset of historical yields was required. In the absence of detailed field measurements at the parcel level, wheat yield data were simulated based on national-level agricultural statistics for Morocco from 2020 to 2024, according to the statistics provided by the Ministry of Agriculture, Maritime Fisheries, Rural Development and Water and Forests. These values, representing annual average wheat yields at the national scale, served as reference points for constructing the simulated dataset.

A statistical simulation strategy was implemented to generate realistic yield values for model training. This approach relied on linking the Normalized Difference Vegetation Index (NDVI) to agricultural yields while preserving the data's distribution structure. For each year, the average NDVI for each parcel was calculated from five monthly observations ($NDVI_1$ to $NDVI_5$). These average NDVI values were then ranked to determine their relative position within the annual distribution using an Empirical Cumulative Distribution Function (ECDF). Each parcel was assigned a score between 0 (lowest NDVI) and 1 (highest NDVI), reflecting its vegetative performance relative to other parcels for the given year. Based on this score, a simulated yield was generated through linear interpolation within a range defined around the annual national mean yield. Parcels with low NDVI values were assigned yields near the minimum, while those with high NDVI values received yields close to the maximum. This process was repeated independently for each year, accounting for year-specific average yields. The resulting dataset provided simulated yields for each parcel and year, offering a coherent and robust training base for machine learning and deep learning models, effectively compensating for the lack of exhaustive field measurements.

2.2.4 Machine Learning models

To predict wheat yields using weather indices and Sentinel-2 imagery, three machine learning models,Random Forest (RF), XGBoost, and Support Vector Machine (SVM) with regression (SVR) were employed. These models were selected for their ability to handle small datasets, robustness to noisy data, and capacity to model non-linear relationships (Aslan et al., 2024). The models were trained using supervised learning, with input features consisting of averaged Sentinel-2 spectral band values (B2, B3, B4, B5, B6, B8) combined with climatic data (precipitation and temperature for November, December of year N-1, and January of year N). Simulated wheat yields served as the target variable.

A multi-horizon approach was implemented exclusively for the model that achieved the best performance, in order to evaluate its prediction capability at different stages of the growing season. The model was trained using cumulative data corresponding to various time windows (January only, January–February, ..., up to January–May), simulating a progressive forecasting ability throughout the agricultural campaign.

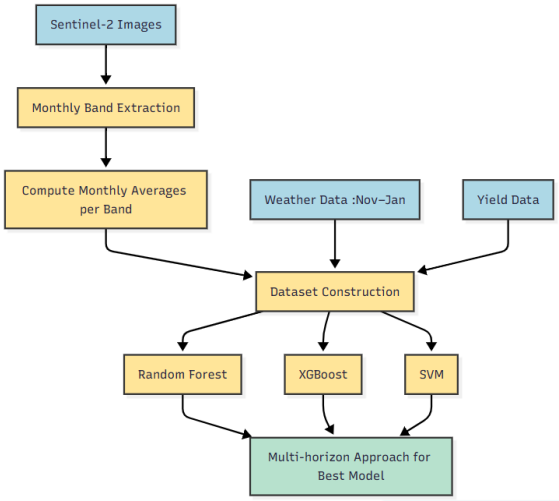

Figure 3. Workflow for Yield Prediction Using Sentinel-2 Images and Machine Learning.

2.2.5 Deep Learning models

2.2.5.1 Our proposed framework STAFNet

We propose a novel framework named STAFNet (Spatial-Temporal Attention Fusion Network) for crop yield prediction. This deep learning model simultaneously integrates sequential multispectral image data and auxiliary tabular data (such as meteorological variables) to capture both the spatio-temporal dynamics of crop growth and relevant environmental factors.

The architecture (Figure 4) is modular and consists of three main components that sequentially process and fuse spatial, temporal, and contextual information:

A Spatial Encoder based on a Vision Transformer (ViT), designed to extract meaningful representations from each individual image in the temporal sequence.

A Temporal and Climate Fusion Module leveraging attention mechanisms (self-attention and cross-attention) to model temporal dependencies and integrate auxiliary data.

A Regression Head, implemented as a Multi-Layer Perceptron (MLP), which predicts the final target value from the fused representation.

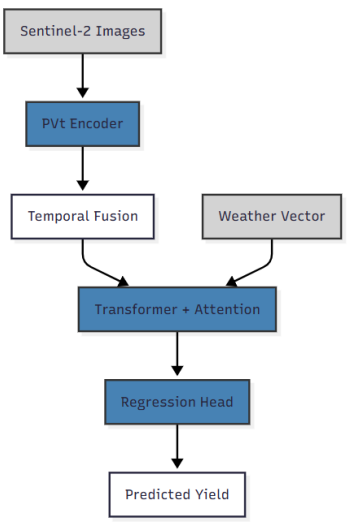

Figure 4. STAFNet components.

(1) Spatial Encoder: Visual Feature Extraction

The first module transforms each multispectral image of shape (C, H, W) in the temporal sequence into a fixed-dimensional feature vector capturing relevant spatial information.

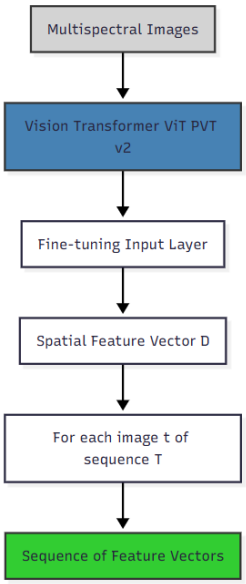

Figure 5. The spatial encoder architecture.

This is achieved using a pretrained Pyramid Vision Transformer v2 (PVT v2) as a feature extractor. Transfer learning allows the model to leverage prior visual knowledge, while only the input projection layer is fine-tuned to handle 6 Sentinel-2 bands; the rest of the network remains frozen. Each input image t in the temporal sequence T is independently encoded into a spatial feature vector, producing a sequence of embeddings. Figure 10 shows the simplified PVT v2 architecture for generating these temporal spatial feature embeddings.

(2) Temporal and Climatic Fusion Module for Spatio-temporal and Contextual Integration

The Temporal and Climate Fusion Module in STAFNet integrates visual features from satellite images with climatic data to generate a comprehensive representation for crop yield prediction. Key components include:

Positional Encoding: Adds temporal order information to the sequence of image embeddings, preserving the crop cycle progression.

Temporal Self-Attention (Transformer): Uses self-attention to capture short- and long-term dependencies between different time steps in the image sequence.

Climate Data Encoding: Climatic variables such as temperature and precipitation are encoded via an MLP into vectors aligned with image embeddings.

Cross-Attention for Fusion: Dynamically integrates temporal image features with climate embeddings, allowing the model to modulate the importance of each based on context.

Final Fusion and Aggregation: Combines temporal and climatic information into a compact feature vector, aggregated over time, which encodes spatial, temporal, and climate information and is passed to the regression head for crop yield prediction.

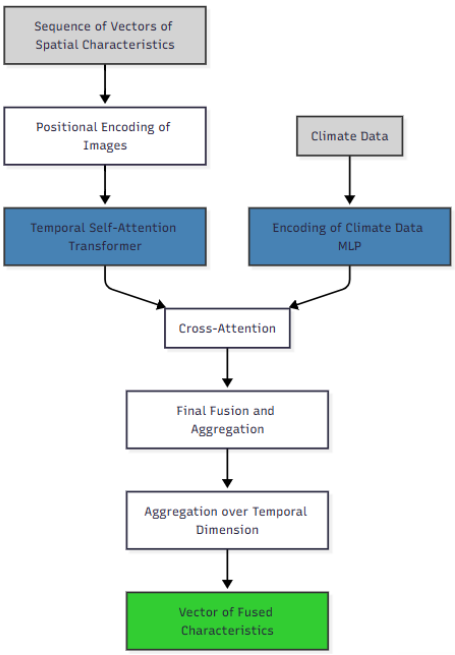

Figure 6. Temporal and Climate Fusion Module.

(3) Regression Head: Final Prediction

The regression head in the STAFNet architecture is responsible for producing the final crop yield prediction based on the fused features extracted from the images and climate data. This component takes as input the integrated representation that combines visual information (from multispectral images), temporal dynamics (from the image sequence), and climatic variables (from meteorological data).The goal of this stage is to transform the fused vector into a scalar value representing the predicted agricultural yield. The regression head is implemented using a **Multi-Layer Perceptron (MLP)**, a simple yet effective neural network composed of several fully connected layers designed for final prediction. Figure 7 illustrates this final phase.

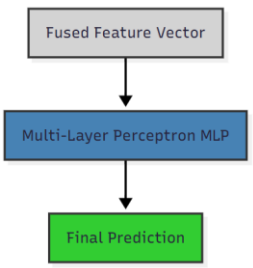

Figure 7. Regression head phase.

STAF-Net was trained using a batch size of 8 for 25 epochs with the AdamW optimizer (initial learning rate 1×10^{-3}) and a cosine annealing learning rate scheduler. The Smooth L1 loss was adopted to improve robustness against noise in ground-truth yield values. The dataset of real field samples was randomly split into 70% training and 30% validation subsets. When GAN-augmented samples were included, they were used exclusively in the training set to enhance diversity without influencing validation performance. To mitigate overfitting, the PVT backbone was largely frozen, weight decay from AdamW was applied, and validation loss was monitored for early stopping. This combination of careful data handling, transfer learning, and regularization enabled stable model training and effective generalization.

2.2.5.2 CNN-LSTM

In parallel with the STAFNet model, another model based on the CNN-LSTM architecture was also tested for crop yield prediction using Sentinel-2 imagery and climate data. The Convolutional Neural Network (CNN) is responsible for extracting spatial features from the images, while the Long Short-Term Memory (LSTM) network captures the temporal dynamics of the climatic variables, such as temperature and precipitation.

The outputs of both modules are fused and passed through a regression layer to estimate crop yields. This model is intended to serve as a baseline for comparison, enabling the evaluation of STAFNet's performance.

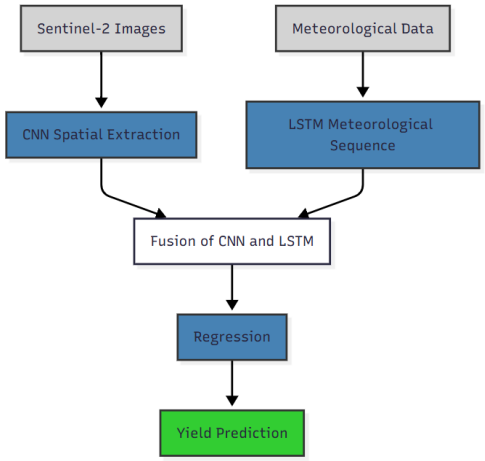

Figure 8. The crop yield prediction process based on the CNN-LSTM architecture.

The same multi-horizon approach was implemented exclusively for the model that achieved the best performance, in order to assess its prediction capability at different stages of the growing season. And to ensure fairness between the two models, STAFNet and CNN-LSTM were trained using the same hyperparameters: a batch size of 8, an initial learning rate set to 1e-3, and a training duration of 30 epochs. The SmoothL1Loss loss function was selected to reduce sensitivity to outliers while maintaining training stability. These settings are appropriate for the available data volume and help prevent early overfitting while ensuring convergence.

2.3 Data augmentation using GAN

To generate synthetic spectral image patches for data augmentation, a Generative Adversarial Network (GAN) was trained using a set of preprocessed Sentinel-2 image patches of wheat plots. The GAN architecture consisted of a generator and a discriminator trained in an adversarial setup, where the generator aims to produce realistic synthetic images, and the discriminator attempts to distinguish between real and generated samples.

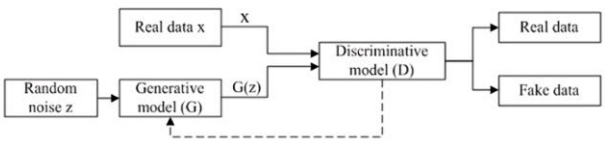

Figure 8. The data generation process based on GANs (Zhang et al.,2022)

The generator receives input noise vectors of dimension 100 sampled from a standard normal distribution and outputs synthetic spectral images of shape **(6, 32, 32)**, corresponding to 6 Sentinel-2 bands. The generator architecture is composed of a sequence of transposed convolutional layers (ConvTranspose2d), each followed by Batch Normalization and ReLU activation functions. The final layer uses a Tanh activation function to normalize pixel values in the range [−1, 1]. The discriminator, on the other hand, is a convolutional neural network designed to classify inputs as real or fake. It consists of a stack of convolutional layers with LeakyReLU activations and Batch Normalization, followed by a Sigmoid output unit to produce a probability score.

During training, the discriminator was updated by minimizing the sum of losses for both real and generated images, while the generator was trained to fool the discriminator by minimizing its ability to detect fake images. This adversarial training was performed iteratively over 50 epochs. At the end of training, the generator was able to produce visually realistic spectral patches, as verified by visual comparison with real Sentinel-2 images. Convergence of the GAN was assessed by monitoring the stabilization of generator and discriminator losses across epochs, supplemented with visual inspection to ensure generated patches resembled real Sentinel-2 images. Once training converged, the generator produced **250 synthetic samples.**

3. Results

This section presents the key findings from the evaluation of various machine learning and deep learning models for wheat yield prediction, highlighting the performance metrics achieved by each approach, particularly focusing on the proposed STAFNet architecture with and without Generative Adversarial Network (GAN) based data augmentation.

3.1 Performance of Classical Machine Learning Models

Among the classical machine learning models evaluated, including Random Forest (RF), Support Vector Machine (SVM), and XGBoost, the XGBoost model demonstrated superior performance. It achieved an R^2 value of 0.919 on the test dataset, indicating a strong correlation between the predicted and actual wheat yields. This result suggests that XGBoost effectively captures the complex, non-linear relationships within the remote sensing and climatic data for accurate yield estimation.

	R^2 (Train)	RMSE (Train)	MAE (Train)	R^2 (Test)	RMSE (Test)	MAE (Test)
RF	0.988	68.60	54.85	0.908	181.54	158.16
XGBoost	0.998	27.47	20.87	0.919	169.54	147.94
SVM	0.838	251.60	189.99	0.703	325.24	268.72

Table 1. Performance of machine learning models on the training and test sets, RMSE and MAE in kg/ha.

To strengthen the evaluation of the XGBoost model, which was identified as the best-performing model during preliminary testing, a k-fold cross-validation (k=5) was conducted to assess the model's stability and generalization capability across the entire dataset. This approach reduces the influence of a single train-test split and provides more representative average performance metrics.

The average performance metrics (R^2, RMSE, MAE) obtained for XGBoost after this cross-validation are presented in the table below:

	R^2 (mean)	RMSE (mean)	MAE (mean)
XGBoost	0.907	186.14	156.91

Table 2. XGBoost average metrics from 5-fold cross-validation.

The following table presents the performance evolution of **XGBoost** across different temporal horizons.

Horizon	Month	R^2	RMSE	MAE
M1	January	0.898	188.43	165.06
M1_M2	February	0.887	198.66	165.88
M1_M2_M3	March	0.897	188.87	163.90
M1_M2_M3_M4	April	0.908	179.26	159.11
M1_M2_M3_M4_M5	May	0.911	176.35	151.75

Table 3. XGBoost performance across temporal horizons.

At the January horizon, the model achieved an R^2 of 0.898, a RMSE of 188.4 kg/ha, and a MAE of 165.1 kg/ha. At the full horizon (M1 to M5), performance increased to an R^2 of 0.911, with a RMSE of 176.3 kg/ha and a MAE of 151.7 kg/ha.

3.2 Performance of Deep Learning Models and the Impact of GAN-based Data Augmentation

The study also evaluated deep learning models, including CNN-LSTM and the newly proposed STAFNet (Spatial-Temporal Attention Fusion Network).

	R^2 (Train)	RMSE (Train)	MAE (Train)	R^2 (Test)	RMSE (Test)	MAE (Test)
STAFNet	0.909	187.95	157.01	0.909	191.50	168.21
CNN-LSTM	0.95	148.40	107.39	0.71	317.00	239.29

Table 4. Performance of deep learning models on training and test sets.

The CNN+LSTM model achieved high performance on the training set, with an R^2 of 0.95, a RMSE of 148.40, and a MAE of 107.39. On the test set, the performance decreased, with an R^2 of 0.71, a RMSE of 317.00, and a MAE of 239.29.
The STAFNet model yielded consistent results between the training and test sets. On the training set, it achieved an R^2 of 0.909, a RMSE of 187.95, and a MAE of 157.01. On the test set, it maintained an identical R^2 of 0.909, with a RMSE of 191.50 and a MAE of 168.21.

A significant aspect of this research was the investigation into the impact of integrating synthetic data generated by GANs for data augmentation on the performance of the STAFNet model.

		R^2	RMSE	MAE
STAFNet -GAN	Train	-	-	-
	Test	0.935	155.93	122.88
STAFNet	Train	0.909	187.95	157.01
	Test	0.909	191.50	168.21

Table 5. STAFNet with vs. without GAN augmentation

The STAFNet model trained with GAN-augmented data achieved an R^2 of 0.935, a RMSE of 155.93 kg/ha, and a MAE of 122.88 kg/ha on the test set.The STAFNet model trained without GAN augmentation obtained an R^2 of 0.909, a RMSE of 191.50 kg/ha, and a MAE of 168.21 kg/ha on the same test set.

During the 5-fold cross-validation of the STAFNet-GAN model, each training fold included both real samples and synthetic images generated by the GAN, while the validation folds consisted exclusively of real samples.
The average performance metrics obtained from this cross-validation were as follows:

	R^2 (mean)	RMSE (mean)	MAE (mean)
STAFNet-GAN	0.90	195.29	165.72

Table 3. STAFNet-GAN average metrics from 5-fold cross-validation.

The following table presents the performance progression of **STAFNet-GAN** across different temporal horizons.

Horizon	Month	R^2	RMSE	MAE
M1	January	0.918	191.19	158.18
M1_M2	February	0.916	202.47	170.30
M1_M2_M3	March	0.918	184.44	160.25
M1_M2_M3_M4	April	0.875	214.84	185.71
M1_M2_M3_M4_M5	May	0.921	192.32	162.88

Table 4. STAFNet-GAN performance across temporal horizons.

4. Discussion

The comparative analysis between classical Machine Learning and Deep Learning models highlights several key findings regarding prediction accuracy, generalization capability, and the role of data augmentation in yield forecasting.
First, among the classical ML approaches, XGBoost emerged as the most accurate, achieving a test R^2 of 0.919, with relatively low RMSE (169.54 kg/ha) and MAE (147.94 kg/ha). Its strong performance was confirmed through 5-fold cross-validation (R^2avg= 0.907), indicating both stability and robustness. Random Forest delivered slightly lower accuracy (R^2 = 0.908) but showed marginally better robustness against overfitting. Conversely, SVM performed poorly (R^2 = 0.703), suggesting

difficulty in modeling the non-linear and high-dimensional structure of the dataset without further feature engineering.
The temporal horizon evaluation demonstrated that predictive accuracy is already high from January onward, with XGBoost achieving $R^2 \approx 0.898$ at M1, and gradually improving to 0.911 by May. This confirms that early-season climatic and satellite information holds substantial predictive value, while additional monthly imagery provides incremental gains.

In the Deep Learning category, the CNN+LSTM model exhibited excellent training performance ($R^2 = 0.95$) but suffered a significant drop on the test set ($R^2 = 0.71$), indicating overfitting and limited generalization. In contrast, STAFNet achieved balanced and stable results between training and test ($R^2 = 0.909$), confirming its ability to capture spatial–temporal dependencies without overfitting. A major finding concerns the integration of GAN-based data augmentation with STAFNet. The STAFNet+GAN configuration outperformed all other tested models, reaching $R^2 = 0.935$ with notably reduced errors (RMSE = 155.93 kg/ha, MAE = 122.88 kg/ha). This improvement was consistent in 5-fold cross-validation ($R^2avg = 0.90$), and in the multi-horizon scenario, where strong accuracy was already achieved in January ($R^2 \approx 0.918$) and maintained until May ($R^2 = 0.921$). The only performance dip occurred in April, possibly due to interannual climatic variability or limitations in the simulated yield targets.

Overall, these results show that while classical ML models like XGBoost remain highly competitive for yield prediction, advanced architectures such as STAFNet especially when enhanced with synthetic data offer superior generalization and stability. The use of GANs proves particularly valuable for enriching the training dataset without additional field campaigns, a significant advantage in agricultural contexts where data acquisition is costly and time-consuming. This synergy between deep learning and synthetic data generation represents a promising avenue for real-time, in-season yield forecasting.

5. Conclusion

This study demonstrated the potential of combining multispectral satellite data, climatic variables, and machine learning models for wheat yield prediction. The results show that traditional Machine Learning approaches, particularly XGBoost, already deliver strong and actionable performance from the early stages of the growing season. However, the STAFNet architecture, enhanced with GAN-based data augmentation, stood out with its superior generalization ability and higher accuracy, outperforming classical methods while maintaining temporal stability.

Synthetic data integration proved to be a strategic lever, enriching training scenarios without additional data collection costs. These findings pave the way for near real-time yield forecasting systems, enabling more responsive and precise agricultural decision-making. Future work may involve extending this methodology to other crops, incorporating additional data sources (phenological indices, soil parameters), and evaluating on real yield data to confirm the robustness of the models under operational conditions.

ACKNOWLEDGEMENT

This work was supported by the Ministry of Agriculture, Fisheries, Rural Development, Water and Forests, Morocco, Program MCRDV (Project CMRM4WP-319/2025-2028).

REFERENCES

Ashfaq, M.; Khan, I.; Alzahrani, A.; Tariq, M.U.; Khan, H.; Ghani, A., 2024. Accurate wheat yield prediction using machine learning and climate-NDVI data fusion. IEEE Access, 12, 40947–40961. https://doi.org/10.1109/ACCESS.2024.3376735

Aslan, M.F.; Sabanci, K.; Aslan, B., 2024. Artificial intelligence techniques in crop yield estimation based on Sentinel-2 data: A comprehensive survey. Sustainability, 16, 8277. https://doi.org/10.3390/su16188277

Ayalew, A.T.; Lohani, T.K., 2023. Prediction of crop yield by support vector machine coupled with deep learning algorithm procedures in Lower Kulfo Watershed of Ethiopia. J. Eng., 2023, 1–8. https://doi.org/10.1155/2023/6675523

Bali, N.; Singla, A., 2021. Deep learning based wheat crop yield prediction model in Punjab region of North India. Appl. Artif. Intell., 35, 1304–1328. https://doi.org/10.1080/08839514.2021.1976091

Bouras, E.H.; Jarlan, L.; Er-Raki, S.; Balaghi, R.; Amazirh, A.; Richard, B.; Khabba, S., 2021. Cereal yield forecasting with satellite drought-based indices, weather data and regional climate indices using machine learning in Morocco. Remote Sens., 13, 3101. https://doi.org/10.3390/rs13163101

Burdett, H.; Wellen, C., 2022. Statistical and machine learning methods for crop yield prediction in the context of precision agriculture. Precis. Agric., 23, 1553–1574. https://doi.org/10.1007/s11119-022-09897-0

Chen, T.; Guestrin, C., 2016. XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794. https://doi.org/10.1145/2939672.2939785

Cunha, R.L.F.; Silva, B., 2020. Estimating crop yields with remote sensing and deep learning. ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci., IV-3/W2-2020, 59–64. https://doi.org/10.5194/isprs-annals-IV-3-W2-2020-59-2020

Dai, T.; Dong, Y., 2020. Introduction of SVM related theory and its application research. In 2020 3rd International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE), Shenzhen, China, 16–18 October 2020; pp. 230–233. https://doi.org/10.1109/AEMCSE50948.2020.00056

Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; Uszkoreit, J.; Houlsby, N., 2021. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929. https://doi.org/10.48550/arXiv.2010.11929

Franz, T.E.; Pokal, S.; Gibson, J.P.; Zhou, Y.; Gholizadeh, H.; Tenorio, F.A.; Rudnick, D.; Heeren, D.; McCabe, M.; Ziliani, M.; Jin, Z.; Guan, K.; Pan, M.; Gates, J.; Wardlow, B., 2020. The role of topography, soil, and remotely sensed vegetation condition towards predicting crop yield. Field Crops Res., 252, 107788. https://doi.org/10.1016/j.fcr.2020.107788

Fu, H.; Lu, J.; Li, J.; Zou, W.; Tang, X.; Ning, X.; Sun, Y., 2025. Winter wheat yield prediction using satellite remote

sensing data and deep learning models. Agronomy, 15, 205. https://doi.org/10.3390/agronomy15010205

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. https://doi.org/10.1016/j.eswa.2021.115511

Genuer, R., n.d. Contributions to random forests methods for several data analysis problems.

Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y., 2020. Generative adversarial networks. Commun. ACM, 63, 139–144. https://doi.org/10.1145/3422622

Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sens. Environ., 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Huber, F.; Yushchenko, A.; Stratmann, B.; Steinhage, V., 2022. Extreme gradient boosting for yield estimation compared with deep learning approaches. Comput. Electron. Agric., 202, 107346. https://doi.org/10.1016/j.compag.2022.107346

Hunt, M.L.; Blackburn, G.A.; Carrasco, L.; Redhead, J.W.; Rowland, C.S., 2019. High resolution wheat yield mapping using Sentinel-2. Remote Sens. Environ., 233, 111410. https://doi.org/10.1016/j.rse.2019.111410

Jabed, M.A.; Azmi Murad, M.A., 2024. Crop yield prediction in agriculture: A comprehensive review of machine learning and deep learning approaches, with insights for future research and sustainability. Heliyon, 10, e40836. https://doi.org/10.1016/j.heliyon.2024.e40836

Khan, M.Y.; Qayoom, A.; Nizami, M.S.; Siddiqui, M.S.; Wasi, S.; Raazi, S.M.K.-R., 2021. Automated prediction of good dictionary examples (GDEX): A comprehensive experiment with distant supervision, machine learning, and word embedding-based deep learning techniques. Complexity, 2021, 2553199. https://doi.org/10.1155/2021/2553199

Krithika, S.; Sangeetha, T.A., 2024. A comprehensive analysis of deep learning models in agricultural crop yield prediction. Int. Res. J. Adv. Eng. Hub IRJAEH, 2, 564–573. https://doi.org/10.47392/IRJAEH.2024.0082

Mancini, A.; Solfanelli, F.; Coviello, L.; Martini, F.M.; Mandolesi, S.; Zanoli, R., 2024. Time series from Sentinel-2 for organic durum wheat yield prediction using functional data analysis and deep learning. Agronomy, 14, 109. https://doi.org/10.3390/agronomy14010109

Pan, Z.; Yu, W.; Yi, X.; Khan, A.; Yuan, F.; Zheng, Y., 2019. Recent progress on generative adversarial networks (GANs): A survey. IEEE Access, 7, 36322–36333. https://doi.org/10.1109/ACCESS.2019.2905015

Pang, A.; Chang, M.W.L.; Chen, Y., 2022. Evaluation of random forests (RF) for regional and local-scale wheat yield prediction in Southeast Australia. Sensors, 22, 717. https://doi.org/10.3390/s22030717

Peng, D.; Cheng, E.; Feng, X.; Hu, J.; Lou, Z.; Zhang, H.; Zhao, B.; Lv, Y.; Peng, H.; Zhang, B., 2024. A deep–learning network for wheat yield prediction combining weather forecasts and remote sensing data. Remote Sens., 16, 3613. https://doi.org/10.3390/rs16193613

Rodrigo, M.; Cuevas, C.; García, N., 2024. Comprehensive comparison between vision transformers and convolutional neural networks for face recognition tasks. Sci. Rep., 14, 21392. https://doi.org/10.1038/s41598-024-72254-w

Schonlau, M.; Zou, R.Y., 2020. The random forest algorithm for statistical learning. Stata J., 20, 3–29. https://doi.org/10.1177/1536867X20909688

Screpnik, C.R.; Zamudio, E.; Gimenez, L.I., 2025. Artificial intelligence in agriculture: A systematic review of crop yield prediction and optimization. IEEE Access, 13, 70691–70697. https://doi.org/10.1109/ACCESS.2025.3560630

Setiawan, D.; Wijayanto, H.; Rahman, L.O.A., 2022. Bagging and random forest classification methods for unbalanced data school dropout cases in Lampung province. In Proceedings of the International Conference on Statistics and Data Science 2021, Bogor, Indonesia, 2021; p. 020026. https://doi.org/10.1063/5.0109130

Sharma, S.; Rai, S.; Krishnan, N.C., 2020. Wheat crop yield prediction using deep LSTM model. arXiv preprint arXiv:2011.01498. https://doi.org/10.48550/arXiv.2011.01498

Shorten, C.; Khoshgoftaar, T.M., 2019. A survey on image data augmentation for deep learning. J. Big Data, 6, 60. https://doi.org/10.1186/s40537-019-0197-0

Srivastava, A.K.; Safaei, N.; Khaki, S.; Lopez, G.; Zeng, W.; Ewert, F.; Gaiser, T.; Rahimi, J., 2022. Winter wheat yield prediction using convolutional neural networks from environmental and phenological data. Sci. Rep., 12, 3215. https://doi.org/10.1038/s41598-022-06249-w

Van Klompenburg, T.; Kassahun, A.; Catal, C., 2020. Crop yield prediction using machine learning: A systematic literature review. Comput. Electron. Agric., 177, 105709. https://doi.org/10.1016/j.compag.2020.105709

Wang, W.; Xie, E.; Li, X.; Fan, D.-P.; Song, K.; Liang, D.; Lu, T.; Luo, P.; Shao, L., 2021. Pyramid Vision Transformer: A versatile backbone for dense prediction without convolutions. arXiv preprint arXiv:2102.12122. https://doi.org/10.48550/arXiv.2102.12122

Wang, Y.; Deng, Y.; Zheng, Y.; Chattopadhyay, P.; Wang, L., 2025. Vision transformers for image classification: A comparative survey. Technologies, 13, 32. https://doi.org/10.3390/technologies13010032

Xue, H.; Yang, Q.; Chen, S.; Ding, W., 2011. SVM: Support vector machines. CRC Press. https://doi.org/10.1201/9781420089653-10

Zhang, J.; Tian, H.; Wang, P.; Tansey, K.; Zhang, S.; Li, H., 2022. Improving wheat yield estimates using data augmentation models and remotely sensed biophysical indices within deep neural networks in the Guanzhong Plain, PR China. Comput. Electron. Agric., 192, 106616. https://doi.org/10.1016/j.compag.2021.106616

Zhao, Y.; Potgieter, A.B.; Zhang, M.; Wu, B.; Hammer, G.L., 2020. Predicting wheat yield at the field scale by combining high-resolution Sentinel-2 satellite imagery and crop modelling. Remote Sens., 12, 1024. https://doi.org/10.3390/rs12061024