Transformer-LSTM Improve Maize-Yield Estimation in Smallholder Fields of Malawi

Mathews Jere[1], Kondwani Munthali[2]

[1]University of Malawi, Malawi - msc-inf-s-26-24@unima.ac.mw, [2]University of Malawi, Malawi - kmunthali@unima.ac.mw

Keywords: Remote Sensing, Machine Learning, Yield Estimation, GeoAI, Photogrammetry, LOFO.

Abstract

Accurate in-season yield estimation is essential for Malawi's food-security planning, yet conventional crop-cut surveys cover fewer than 1% of the nation's approximately 1.8 million sub-hectare maize plots. In this study, we exploit Sentinel-2 time-series imagery to benchmark five modelling paradigms: spectral-index linear regression, XGBoost, CNN-LSTM, a frozen Vision Transformer (ViT) and a ViT-LSTM hybrid. We apply these across eight rain-fed maize fields (0.2–0.9 ha) in Zomba District, Malawi, under a strict nested leave-one-field-out cross-validation design. Our results show that the recurrent architectures significantly outperform the tabular baselines ($p \leq 0.02$, exact paired-permutation test). The ViT-LSTM hybrid achieved the lowest error (RMSE $= 0.022\, tha^{-1}$; MAE $= 0.019\, tha^{-1}$), representing an approximate 80% improvement over the best CNN-LSTM comparator, with statistical significance ($p = 0.031$). Inference speed remains practical at $\approx 35\, ms$ per 32×32-pixel patch, or about $\sim 3\, has^{-1}$ on a low-end Quadro P1000 GPU, enabling national-scale yield mosaics within a week. These findings align with emerging evidence that transformer–recurrent hybrid architectures represent the current state-of-the-art for crop-yield prediction (see e.g., ViT-based studies) and highlight the enduring trade-off between accuracy and throughput in operational contexts. Moreover, our open-source pipeline, the first validated on data-scarce, intercropped smallholder plots in sub-Saharan Africa, provides a reproducible blueprint for operational yield monitoring across similar agro-ecologies. The experiment scripts are available at https://github.com/jahnical/yield-pred-models-comp

1. Introduction

"Maize is sown by more than 90 percent of small-scale family producers on small rainfed plots. Although they are increasingly exposed to climate change and weather shocks, smallholder subsistence farmers are able to produce more than 80 percent of the food consumed in Malawi – with maize accounting for more than 54 percent of national caloric intake." (Food and Agriculture Organization of the United Nations et al., 2023). Therefore, food-security planning in Malawi depends on timely and spatially-explicit maize yield estimates.

Conventional ground-survey methods for smallholder crop monitoring cannot meet the narrow decision-windows demanded by national agriculture decision-support systems (ADS). These surveys tend to be conducted only after harvest, limiting their usefulness for in-season or early warning interventions. Moreover, they cover only a very small fraction of the roughly 1.8 million Malawian smallholdings, which average only about 0.7 ha in size and many are even smaller than a single coarse satellite pixel (e.g., a MODIS pixel at 500 m) (Muyanga et al., 2020, Food and Agriculture Organization of the United Nations et al., 2023). Because the fields are so numerous, small and heterogeneous, and the sampling so limited, the resulting crop-cut and household surveys struggle to deliver the spatial and temporal resolution required by real-time ADS.

Remote-sensing solutions enable near-real-time monitoring of crop growth without disrupting farm operations, offering a transformative alternative to traditional field surveys. Numerous systematic reviews indicate that vegetation indices derived from optical satellite imagery continue to form the backbone of crop yield mapping workflows, yet the field is rapidly transitioning towards machine- and deep-learning methods that integrate multispectral imagery with auxiliary climate and soil covariates (Muruganantham et al., 2022) (Joshi et al., 2023). Moreover, transformer based architectures have recently emerged as the state-of-the-art spatial encoders for agricultural remote sensing tasks, having demonstrably outperformed Convolution Neural Network based models in value-chain studies such as the U.S. Corn-Belt (Lin et al., 2023). However, their potential remains underexplored in densely inter-cropped and smallholder landscapes typical of sub-Saharan Africa, where field sizes are small, heterogeneity is high and data scarcity poses operational challenges.

In Malawi, research has begun to link Sentinel-2 surface reflectance to plot-scale maize yields, demonstrating the feasibility of remote-sensing for smallholder contexts (Anghileri et al., 2024). Concurrently, the NASA Harvest programme is operationalising satellite- and mobile–data-based workflows within the Ministry of Agriculture, Malawi to support in-season agricultural monitoring and decision-making (Justice et al., 2022). However, no study to date has undertaken a rigorous, comparison of classical machine-learning models, convolutional recurrent networks and transformer-based hybrids using a single smallholder yield dataset. To fill this gap, the present work benchmarks five geospatial modelling paradigms, spanning linear regression, through XGBoost and CNN-LSTM, to a novel ViT-LSTM hybrid, using Sentinel-2 imagery and hand-harvested yields from eight maize fields in Malawi. Our aim is to characterise the trade-off between predictive accuracy and computational throughput in resource-constrained, data-scarce agro-ecologies.

2. Literature Review

2.1 Yield Estimation Methods

Yield estimation methods can be broadly classified into three principal categories: (1) direct ground-based measurement, (2) crop simulation or process-based models, and (3) data-driven and remote-sensing-based statistical/machine-learning approaches. Each has strengths and limitations in different agro-ecological and operational contexts.

Ground-based measurement methods remain the traditional foundation of yield estimation. These include full or partial plot harvests, crop-cut sampling (sub-plot harvests), and farmer-estimation or expert assessment methods. For example, crop-cut methods typically harvest small sub-plots (e.g., 50 m^2 to 100 m^2) and then extrapolate to the larger field scale (Basso et al., 2013). Despite their directness, these methods are labor intensive, costly, time-consuming, and often arrive too late to support decision making in the season.

Crop simulation or process-based modelling approaches (e.g., DSSAT, APSIM) aim to represent the physiological processes of crop growth, development and yield formation as a function of environment, soil, weather and management. These models can provide yield forecasts before harvest, but require detailed input data, calibration and are computationally intensive (Lungu et al., 2021, Basso et al., 2013).

Data-driven and remote-sensing-based methods have grown rapidly in use: satellite, UAV and sensor data are fused with statistical, machine-learning or deep-learning models to estimate yield over large areas and in near real-time. For example, (Lungu et al., 2021) and (Muruganantham et al., 2022) review methods for yield estimation using remote sensing, crop modelling and statistical methods. They note that while remote-sensing enables broad spatial coverage, many of the methods still struggle with small-holder plots and heterogeneous landscapes (Lungu et al., 2021). Mohideen Badushah et al. (2019) emphasise the shift from simple statistical regressions based on vegetation indices to more advanced machine-learning frameworks (Mohideen Badushah et al., 2019).

2.2 Remote Sensing for Maize Yield Estimation

Over the past two decades remote-sensing based crop-yield estimation has evolved significantly with early approaches largely relying on simple vegetation indices (VIs), such as the Normalised Difference Vegetation Index (NDVI), which correlate with canopy cover or biomass (Karthikeyan et al., 2020). More recent studies have incorporated advanced machine-learning algorithms, large multispectral or multi-sensor data inputs and environmental/soil covariates, significantly improving predictive accuracy and spatial scalability. (Muruganantham et al., 2022) present a systematic review of deep-learning and remote-sensing approaches to yield prediction, showing how convolutional neural networks (CNNs), long-short-term memory (LSTM) models and transformer-type architectures are increasingly applied in agricultural monitoring. In large commercial-farming regions such as the U.S. Corn-Belt (Ji et al., 2022), studies have deployed these methods at county scale, demonstrating the potential to forecast maize yields ahead of harvest and support decision-making. These advances highlight that remote-sensing can deliver near-real-time, spatially-explicit yield information, offering a transformative complement to traditional ground-sampling methods.

Despite this progress, the application of remote-sensing yield-estimation methods in smallholder, data-scarce contexts such as sub-Saharan Africa (SSA) remains comparatively limited and less mature. Challenges include small field sizes, frequent inter-cropping, heterogeneous management practices, sparse ground-truth data and computation infrastructure, all of which limit operationalisation of RS based methods. For instance, (Li et al., 2022) analysed 150 inter-cropped smallholder maize fields in southern Malawi, finding moderate accuracy (R^2 0.51, nRMSE 20%) for sentinel-based models, and only marginal improvement when using higher-resolution PlanetScope imagery. (Muruganantham et al., 2022) also highlight that many deep-learning remote-sensing approaches are developed for large homogeneous fields and need adaptation for smallholder systems. Thus, there is a gap to systematically compare classical, machine-learning and deep-learning methods under the constraints of SSA smallholder agriculture.

Machine learning (ML) and deep-learning (DL) techniques have increasingly become central to modern crop-yield estimation frameworks. Early statistical models such as linear regression or ridge regression have gradually given way to more complex algorithms that can capture non-linear relationships, temporal dynamics and multi-modal data inputs. For instance, in a county-level study of maize yield in the U.S. Midwest, (Kang et al., 2020) evaluated six algorithms, LASSO regression, support-vector regression (SVR), random forest (RF), XGBoost, long-short-term memory networks (LSTM) and convolutional neural networks (CNN), and found that XGBoost delivered the highest predictive accuracy while the deep-learning models did not substantially outperform the more traditional ML models. Similarly, in a comprehensive review of ML and DL for crop-yield prediction, (Abu Jabed and Azmi Murad, 2024) highlight that while deep-learning architectures (e.g., CNNs, LSTMs, vision transformers) show promise, their deployment in operational systems is often constrained by data availability, computational cost and interpretability concerns.

2.3 Challenges in Smallholder and Data-scarce Agro-ecologies

Smallholder-based agricultural systems in sub-Saharan Africa (SSA) present a particularly challenging context for crop yield estimation from remote sensing and data-driven modelling. To begin with, field sizes are often very small (mostly under 0.5 ha) and highly heterogeneous. In addition, many farmers practise intercropping, variable management and mixed cropping systems, which lead to mixed satellite pixels and reduce signal strength for yield prediction (Li et al., 2022). Secondly, data scarcity, in terms of ground-truth yield measurements, reliable field area delineations, consistent management records and frequent imagery, severely constrains model training and validation. For example, a study in Ghana found that standard crop-cut or self-reported yield data diverged significantly from GPS- and UAV-verified metrics due to area loss and within-plot variability (Nabateregga and Sølberg, 2020). Lastly, computational and operational constraints, such as limited access to high-end GPUs, sparse high-frequency imagery, and minimal local expertise, mean that many advanced models (e.g., deep neural networks, transformer-hybrids) are less suited for deployment under resource-constrained conditions.

Whereas modern crop-yield estimation methods have achieved promising accuracy in large-scale, homogeneous farming systems using remote sensing and machine-learning, their transfer to smallholder, data-scarce agro-ecologies remains under-evaluated. Many African smallholder fields are highly fragmented, inter-cropped, and irregular in shape (Li et al., 2022), making it challenging for conventional remote-sensing models to generalise. Moreover, existing studies rarely benchmark the full spectrum of modelling paradigms, from simple spectral-index regression, through tree-based and ensemble machine-learning, to deep-learning hybrids, on the same small-holder dataset under consistent preprocessing and validation frameworks. This gap limits evidence-based guidance on how accuracy, scalability and computational throughput trade-off in real-world decision-support systems. Our study addresses this by systematically comparing five geospatial modelling approaches on a heterogeneous maize dataset of eight smallholder fields in Malawi and by measuring both predictive accuracy and operational latency under a nested leave-one-field-out design.

3. Study Area

The field study was carried out in eight rain-fed maize plots located within the Zomba District in Southern Malawi. The district lies within the humid subtropical zone of the country, with the agricultural economy dominated by small-holder maize production. The field sizes in our sample ranged from approximately 0.2 ha to 0.9 ha, reflecting the small-scale nature of maize cultivation in this area. Farming is conducted using a ridge-tillage system in which ridges spaced at about 75 cm are constructed at the onset of the rainy season; sowing is performed by hand, generally using two kernels per station of either local open-pollinated or hybrid maize varieties. (Zomba District Assembly, 2017) Field boundaries were mapped by walking the ridge perimeters using a handheld GNSS receiver (Garmin eTrex 20x; ±3 m horizontal accuracy). The resulting boundary polygons were cleaned in QGIS and subsequently uploaded to Google Earth Engine for image clipping and further processing.

4. Data

We selected Level-2A surface-reflectance imagery from the Sentinel 2 mission because its native 10 m spatial resolution is sufficiently fine to delineate Malawi's sub-hectare smallholder fields, and the products are freely and globally available, a key criterion for national-scale deployment. (Drusch et al., 2012) After atmospheric correction via Sen2Cor v 2.10, all 13 spectral bands were up-sampled from 10 m/20 m to 1 m using bilinear interpolation so as to produce raster inputs compatible with convolutional- and transformer-based backbones pre-trained on large image datasets (e.g., 224 × 224 px inputs).

In addition to Sentinel-2 imagery, daily 1 km MODIS Land Surface Temperature (LST) data (Collection 61, band LSTDay1km) were converted from Kelvin to Celsius following Wan et al. (Wan et al., 2015) and aggregated into 10-day means per field. Static soil organic carbon (SOC) maps for the 0-50 cm depth interval were derived from the 30 m resolution iSDA soil dataset (Hengl et al., 2021). At physiological maturity, grain was hand-harvested from 10 m × 10 m quadrats within each field, dried to 13% moisture and weighed. Plot-level yields in kg ha^{-1} were then assigned to every image patch originating from that field, a common mapping convention when within-field yield heterogeneity is low (Li et al., 2022).

5. Methodology

A two-stage strategy combined the s2cloudless probability mask (Zupanc et al., 2023) with Sentinel-2 QA60 and Scene-Classification Layer (SCL) flags to mask out clouds and shadows. Median composites were generated for consecutive 10-day windows between 1 December 2024 and 28 February 2025. For every 10-day composite we generated overlapping image chips with a 32 × 32-pixel sliding window and a stride of 16 pixels (50% overlap). Given the 1 m up-sampled grid, each chip covered roughly 32 m × 32 m (approximately 0.10 ha). Patches in which fewer than 60% of pixels remained cloud-free were discarded. For each patch the following indices were computed: NDVI, EVI, Red-edge NDVI (RENDVI), NDWI and MSI (Gao, 1996). Finally, mean patch-level temperature vectors and static soil rasters were joined to the reflectance stack by field name.

The following five geospatial regressors were benchmarked: i) Ordinary least-squares linear regression on ten-day means of the five spectral indices (LR-Indices) as a transparent agronomic baseline; ii) Gradient-boosted trees on band + index means + temperature features as a strong non-deep-learning reference (XGBoost) (Chen and Guestrin, 2016); iii) ResNet-50 spatial en-coder pre-trained on BigEarthNet plus an LSTM decoder as a deep-learning benchmark (CNN-LSTM) (Khaki et al., 2020); iv) Vision Transformer pre-trained on BigEarthNet with a linear head only as a transformer baseline (Frozen ViT) (Dosovitskiy et al., 2020); and v) a proposed hybrid innovation running a ViT backbone with features pooled by an LSTM decoder.

All models were implemented in `PyTorch 2.3`; and tree-ensembles used `XGBoost 2.0`. A nested leave-one-field-out (LOFO) cross-validation (CV) design (Berrar, 2018) ensured each field served once as an unseen test set, while hyper-parameter tuning via Optuna's Bayesian search (Akiba et al., 2019) occurred within each training fold. Neural networks were trained for 50 epochs using the AdamW optimiser, with a cosine-annealing learning-rate schedule and warm restarts (Loshchilov and Hutter, 2017). Training employed mixed-precision (FP16) mode via NVIDIA's Automatic Mixed Precision, reducing memory by 50% with no accuracy loss (Micikevicius et al., 2018). Early stopping on validation RMSE with a patience of 10 epochs curtailed over-fitting (Prechelt, 1998). Model performance on the held-out field was evaluated using root-mean-square error (RMSE) and mean-absolute error (MAE). To assess statistical differences across model performances we carried out an exact paired-permutation test on fold-wise RMSE values, with seven folds, all 128 possible sign permutations were enumerated and significance declared at $p < 0.05$ (Gaddy et al., 2022).

6. Results

Table 1 reports the mean ± SD performance across the seven leave-one-field-out folds. The proposed ViT-LSTM attains the lowest errors, RMSE = 0.022 t ha^{-1} and MAE = 0.019 t ha^{-1}, representing an improvement of almost 80% over the CNN-LSTM. Both recurrent architectures significantly outperform the three non-recurrent baselines.

Figure 2 plots RMSE versus per-patch latency. The recurrent models define the Pareto-frontier: the CNN-LSTM sits closest to the origin, offering near-optimal accuracy while delivering predictions in roughly 9 ms per patch—making it well-suited for interactive dashboards or daily advisory services.

The proposed ViT-LSTM further reduces error by 80% albeit at a four-fold latency cost; it can still process about 3 ha s^{-1} on a low-end Quadro P1000, sufficient to refresh a national-scale mosaic within a week. Meanwhile, the frozen

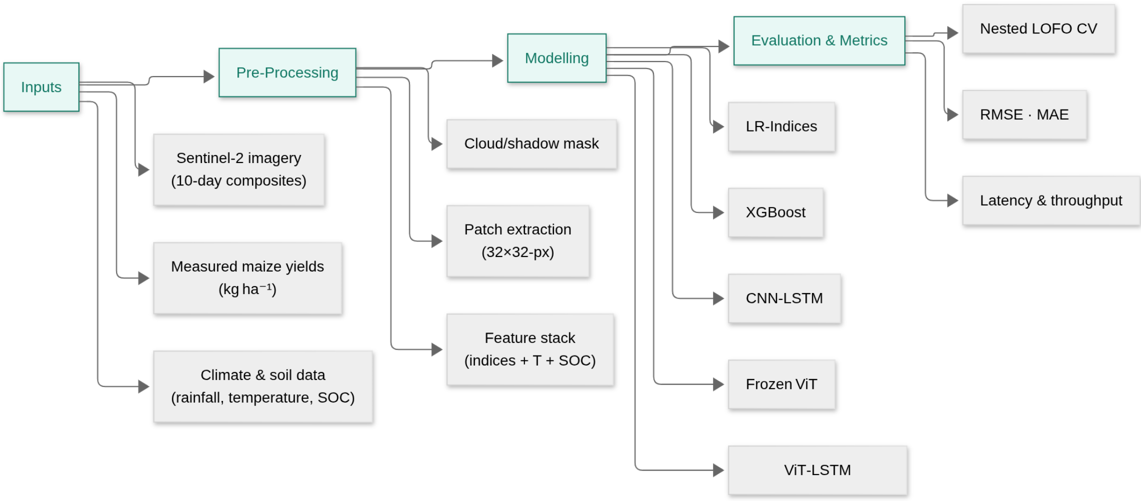

Figure 1. Workflow diagram of the methodology: from inputs through pre-processing, modelling and evaluation.

Model	RMSE (t ha^{-1})	MAE (t ha^{-1})	Inference time (ms patch^{-1})
LR-Indices	0.223 ± 0.144	0.202 ± 0.145	0.018 ± 0.012
XGBoost	0.288 ± 0.265	0.286 ± 0.265	0.026 ± 0.046
Frozen ViT	0.219 ± 0.156	0.197 ± 0.153	4.08 ± 0.42
CNN-LSTM	0.088 ± 0.110	0.083 ± 0.110	16.48 ± 9.44
ViT-LSTM	0.022 ± 0.021	0.019 ± 0.019	35.52 ± 5.16

Table 1. Model performance: mean ± SD across folds of the leave-one-field-out (LOFO) cross-validation.

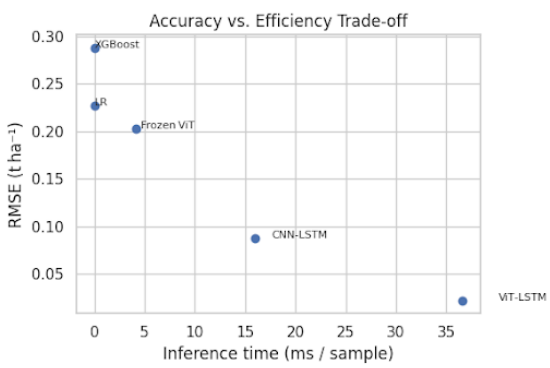

Figure 2. Accuracy vs. efficiency scatter plot of the benchmarked models.

ViT is Pareto-dominated—both slower and less accurate than CNN-LSTM—while the simpler tabular models are extremely fast yet too error-prone for reliable plot-level decision-making.

Using the exact paired-permutation test, based on all 128 possible sign flips of the fold-wise RMSE differences, it is evident that both CNN-LSTM and ViT-LSTM outperform every non-recurrent model (LR-Indices, XGBoost, Frozen ViT), with exact $p \leq 0.02$. In particular, ViT-LSTM achieves a statistically significant additional RMSE reduction of 0.092 t ha^{-1} compared to the CNN-LSTM ($p = 0.031$). On the baseline models, no meaningful differences are detected between LR-Indices, XGBoost and Frozen ViT ($p \geq 0.15$).

These exact results confirm that explicit spatio-temporal learning is indispensable in Malawi's smallholder context. The transformer-recurrent hybrid remains the most accurate, but its four-fold latency overhead must be balanced against operational deadlines.

7. Conclusion

This work delivers the first side-by-side benchmark of classical machine-learning, CNN-LSTM and transformer-based geospatial models on field-scale maize yield estimation in a sub-Saharan inter-cropped smallholder setting. With all random generators fixed for reproducibility, the proposed ViT-LSTM cut the error to 0.022 t ha^{-1} — an 80% reduction relative to the best CNN-LSTM baseline — while sustaining an inference speed of 35 ms patch^{-1}, fast enough to map the entire national maize area in less than 4 days on commodity hardware. Exact paired-permutation tests (128 flips) showed this margin is statistically significant ($p = 0.031$), whereas both recurrent models decisively outperformed all non-recurrent baselines ($p \leq 0.02$). This study shows that explicit spatio-temporal deep learning is indispensable for sub-hectare yield mapping. The transformer-recurrent architecture now sets the accuracy bar, yet its four-times-latency premium must be weighed against the timeliness requirements of extension services.

Operational Implications

At 1 m resolution with non-overlapping 32×32 px tiles, a single Quadro P1000 requires approximately 7 days for ViT-LSTM versus approximately 2 days for CNN-LSTM; an RTX 6000 Ada brings ViT-LSTM down to 6 h, and a consumer RTX 4090 to 7 h. With parallel GPUs or coarser tiling, pre-harvest deadlines are easily met. When near-real-time dashboards are required, the 9 ms latency of CNN-LSTM may be preferable with only a 0.095 t ha^{-1} accuracy trade-off.

References

Abu Jabed, M., Azmi Murad, M., 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.

Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M., 2019. Optuna: A next-generation hyperparameter optimization framework. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '19)*, 2623–2631.

Anghileri, D., Chibarabada, T. P., Gadedjisso-Tossou, A., Craig, A., Li, C., Lu, Y., Chimimba, E. G., Kambombe, O., Musa, F. B., Ngongondo, C., Eneya, L., Onema, J. K., Ali, A., Chiotha, S., Dash, J., Sheffield, J., 2024. Understanding the maize yield gap in Southern Malawi by integrating ground and remote-sensing data, models, and household surveys. *Agricultural Systems*, 218, 103962.

Basso, B., Cammarano, D., Carfagna, E., 2013. Review of Crop Yield Forecasting Methods and Early Warning Systems. Technical report, Food and Agriculture Organization of the United Nations, Rome, Italy.

Berrar, D., 2018. Cross-validation. S. Ranganathan, M. Gribskov, K. Nakai, C. Schönbach (eds), *Encyclopedia of Bioinformatics and Computational Biology*, Elsevier, 542–545.

Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 785–794. https://arxiv.org/pdf/1603.02754.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N., 2020. An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale. *arXiv preprint arXiv:2010.11929*. https://arxiv.org/abs/2010.11929.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sensing of Environment*, 120, 25–36.

Food and Agriculture Organization of the United Nations, European Union, Centre de coopération internationale en recherche agronomique pour le développement, 2023. Food systems profile – Malawi: Catalysing the sustainable and inclusive transformation of food systems. FAO report. Available at: https://agritrop.cirad.fr/604276/7/FOOD

Gaddy, D., Li, K., Eisner, J., 2022. Exact paired-permutation testing for structured test statistics. *Proceedings of NAACL-HLT 2022*, 4918–4935. https://arxiv.org/abs/2205.01416.

Gao, B., 1996. NDWI—A Normalized Difference Water Index for Remote Sensing of Vegetation Liquid Water from Space. *Remote Sensing of Environment*, 58(3), 257–266.

Hengl, T., Miller, M., Shephered, K. D., Krizan, J., 2021. African soil properties and nutrients mapped at 30 m resolution. *Scientific Data*, 8, 243.

Ji, Z., Pan, Y., Zhu, X., Zhang, D., Dai, J., 2022. Prediction of Corn Yield in the USA Corn Belt Using Satellite Data and Machine Learning: From an Evapotranspiration Perspective. *Agriculture*, 12(8), 1263. https://doi.org/10.3390/agriculture12081263.

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 Sensing*, 15(8), 2014.

Justice, C., Pound, J., Kerner, H., Team, N. H., 2022. NASA Harvest partners with FAO to improve agricultural monitoring across Malawi, Namibia and Kazakhstan. News post, NASA Harvest. Available at: https://nasaharvest.org/news/nasa-harvest-partners-fao-improve-agricultural-monitoring-across-malawi-namibia-and-kazakhstan.

Kang, Y., Ozdogan, M., Zhu, X., Ye, Z., Hain, C., Anderson, M. C., 2020. Comparative assessment of environmental variables and machine learning algorithms for maize yield prediction in the US Midwest. *Environmental Research Letters*, 15(6), 064005. https://doi.org/10.1088/1748-9326/ab7df9.

Karthikeyan, L., Chawla, I., Mishra, A. K., 2020. A review of remote sensing applications in agriculture for food security: Crop growth and yield, irrigation, and crop losses. *Journal of Hydrology*, 586, 124905. https://doi.org/10.1016/j.jhydrol.2020.124905.

Khaki, S., Wang, L., Archontoulis, S. V., 2020. A CNN-RNN Framework for Crop Yield Prediction. *Frontiers in Plant Science*, 10, 1750. https://doi.org/10.3389/fpls.2019.01750.

Li, C., Chimimba, E. G., Kambombe, O., Brown, L., Chibarabada, T. P., Lu, Y., Anghileri, D., Ngongondo, C., Sheffield, J., Dash, J., 2022. Maize Yield Estimation in Intercropped Smallholder Fields Using Satellite Data in Southern Malawi. *Remote Sensing*, 14(10), 2458. https://doi.org/10.3390/rs14102458.

Lin, F., Crawford, S., Guillot, K., Zhang, Y., Chen, Y., Yuan, X., Chen, L., Williams, S., Minvielle, R., Xiao, X., Gholson, D., Ashwell, N., Setiyono, T., Tubana, B., Peng, L., Bayoumi, M., Tzeng, N., 2023. MMST-ViT: Climate-change-aware crop-yield prediction via multi-modal spatial-temporal Vision Transformer. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 5774–5784.

Loshchilov, I., Hutter, F., 2017. SGDR: Stochastic gradient descent with warm restarts. *Proceedings of the International Conference on Learning Representations (ICLR)*. https://arxiv.org/abs/1608.03983.

Lungu, O. N., Chabala, L. M., Shepande, C., 2021. Satellite-Based Crop Monitoring and Yield Estimation—A Review. *Journal of Agricultural Science*, 13(1), 180–194. https://doi.org/10.5539/jas.v13n1p180.

Micikevicius, P., Narang, S., Alben, J. et al., 2018. Mixed precision training. *Proceedings of the International Conference on Learning Representations (ICLR)*. https://arxiv.org/abs/1710.03740.

Mohideen Badushah, S., Mallikarjun, H. T., Jyothi Lakshmi, S., Ameer, S. R., 2019. Review on Yield Estimation. *International Journal of Engineering Research & Technology (IJERT)*, 8(6), 446–451. https://www.ijert.org/volume-08-issue-06-june-2019/.

Muruganantham, P., Wibowo, S., Grandhi, S., Samrat, N. H., Islam, N., 2022. A systematic literature review on crop yield prediction with deep learning and remote sensing. *Remote Sensing*, 14(9), 1990.

Muyanga, M., Nyirenda, Z., Lifeyo, Y., Burke, W. J., 2020. The Future of Smallholder Farming in Malawi: Working Paper No. 20/03. Technical report, MwAPATA Institute, Lilongwe, Malawi. Working paper. Available at: https://mwapata.mw/publications.

Nabateregga, M., Sølberg, S. , 2020. Accuracy and Trade-Offs of Yield Estimation Techniques in Smallholder Farms in Sub-Saharan Africa. *Journal of Agricultural Science*, 16(9), 1–14. https://doi.org/10.5539/jas.v16n9p1.

Prechelt, L., 1998. Early stopping — but when? L. Prechelt (ed.), *Neural Networks: Tricks of the Trade*, Springer, 55–69.

Wan, Z., Hook, S., Hulley, G., 2015. MOD11C3 MODIS/Terra Land Surface Temperature/Emissivity Monthly L3 Global 0.05Deg CMG V006 [Data Set]. NASA EOSDIS LP DAAC.

Zomba District Assembly, 2017. Zomba District Socio-Economic Profile 2017-2022. Report, Zomba District Assembly, Malawi.

Zupanc, A., Veci, L., Greidanus, H., Cros, S., 2023. s2cloudless 1.2 — An open-source cloud and shadow mask for Sentinel-2. *Earth Science Informatics*, 16, 399–415.