Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Linking physiological nitrogen responses to spectral signatures: a multi-spectral imagery dataset (RGB and NDVI) for nitrogen deficiency detection in rice (Oryza sativa) across growth stages

Lwekiza Nelson Ndiwaita

lwekizandiwaita@gmail.com

Sokoine University of Agriculture

Kadeghe Goodluck Fue

Sokoine University of Agriculture

Mawazo J Shitindi

Sokoine University of Agriculture

Research Article

Keywords: Rice, Nitrogen deficiency, NDVI, RGB imagery, Image dataset, Precision agriculture

Posted Date: November 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11284031/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Open-access imagery datasets are essential for training, validating and comparing machine-learning models in agriculture, yet none exists for nitrogen (N) deficiency in rice grown by smallholder farmers in East Africa. This study produced paired Red-Green-Blue (RGB) and Normalised Difference Vegetation Index (NDVI) images of rice leaves showing N deficiency symptoms at six growth stages (40–90 days after planting, DAP). A field experiment with a randomised complete block design (four replications) tested four N rates: 0 (T0), 50 (T1), 100 (T2) and 150 kg N ha^{-1} (T3), using the variety SARO 5 (TXD306) in Morogoro, Tanzania. Images were captured with a GoPro Hero 10 (RGB) and a Mapir Survey3 (NDVI) camera. From 12,000 raw images, automated and manual quality control retained 8,000 images (4,000 RGB and 4,000 NDVI; 1,000 per class). Plant height, leaf area, leaf tissue N and grain yield were analysed by one-way analysis of variance (ANOVA) with Tukey's honestly significant difference (HSD) test ($P \leq 0.05$). All four variables increased with N rate; T2 and T3 did not differ at most sampling dates, while T0 and T1 showed clear deficiency. Mean NDVI was strongly correlated with laboratory-measured tissue N ($r = 0.893$, $R^2 = 0.798$, $P < 0.001$), and a baseline MobileNetV3 classifier reached 75.3% accuracy in distinguishing the four N classes from RGB images. The openly available dataset links visible symptoms to measured N status and provides a foundation for deep-learning tools for N management in smallholder rice systems.

1. Introduction

Rice (*Oryza sativa*) supplies about 20% of the world's dietary energy, compared with 19% for wheat and 5% for maize (Wanasundara et al. 2024). Rice production in Tanzania has grown from about 94,000 t in the 1960s to about 1.8 million t in the 2010s and is expected to keep rising with population demand (Alfred et al. 2021). Morogoro Region is the largest paddy producer in mainland Tanzania (572,884 t; 17.2% of national output), followed by Mbeya (413,212 t; 12.4%) and Shinyanga (381,092 t; 11.4%) (Ministry of Agriculture 2021).

Rice yields can be raised through improved varieties, judicious fertiliser use, irrigation and good crop management (Shrestha et al. 2022). Nitrogen (N) is the nutrient most often limiting in rice. It supports photosynthesis, tillering, leaf area development, grain formation, grain filling and protein synthesis, and demand is greatest during early and mid-tillering, panicle initiation, booting and grain development (Christiaan Struik et al. 2022; Shrestha et al. 2022). Modern production relies heavily on synthetic N fertiliser, which has increased yields but harms water, air and ecosystems when over-applied (Kumar et al. 2021; Liu et al. 2025; Savci 2012; Skiba and Rees 2015). Farmers often apply large amounts of N to maximise yield, yet only 20–50% is taken up by the crop (Coggins et al. 2025; Vlek and Byrnes 1986), and efficient N use remains a challenge in irrigated lowland, rain-fed lowland and upland rice systems (Aulakh 1996; Sekiya et al. 2025).

In Tanzania, traditional farming practices, minimal fertiliser use, weak fertiliser markets, limited policy and institutional support and ineffective extension services have contributed to nutrient depletion, low

yields and poverty (Aworka et al. 2022; Chianu et al. 2012; Sithole and Olorunfemi 2024). The average yield of improved varieties such as SARO 5 remains far below their genetic potential of about 10 t ha^{-1} because of nutrient deficiencies, poor nutrient management and poor soil-water management (Sutton 2015). Conventional diagnosis of N deficiency relies on plant sampling and chemical analysis, which is labour-intensive, slow and costly (Kavishe 2021). Farmers therefore need affordable, timely decision-support tools (Aworka et al. 2022).

Sensor-based precision agriculture can provide timely, spatially resolved measurements of crop biophysical parameters to guide N management (Fue et al. 2025; Karunathilake et al. 2023; Mgohele et al. 2025; Sharma and Shivandu 2024). Images of N deficiency symptoms can be used to train machine-learning models that run in mobile applications and help extension officers and farmers detect crop stress early (Laizer et al. 2024; Sharma et al. 2022). Artificial intelligence is already improving N use efficiency in many regions (Akintuyi 2024), but progress in Tanzania is constrained by limited documentation and data, skills, infrastructure and the technology ecosystem (Ly 2021). Open imagery datasets for Tanzanian crops are beginning to appear (Laizer et al. 2024; Mduma and Laizer 2023), but we are not aware of an open dataset for N deficiency in rice.

Two low-cost imaging modalities are well suited to this task. The Normalised Difference Vegetation Index (NDVI), computed from near-infrared (NIR) and red reflectance, is widely used to monitor vegetation status and cover (Jiang et al. 2021; Li et al. 2021). RGB imagery offers high-resolution information on leaf colour and texture that changes with N status (Zheng et al. 2018). The objective of this study was to build and describe an open, paired RGB and NDVI image dataset of rice leaves grown under four N rates across six growth stages, to link the images to measured agronomic and tissue N data, and to verify their usefulness for developing machine-learning models of N deficiency in smallholder rice systems.

2. Materials and methods

2.1. Study site

The experiment was carried out at a university research farm (crop museum) in Morogoro Region, Tanzania (7°25' S, 38°04' E; 540 m above sea level). The site has a mean annual air temperature of 24°C and a mean annual rainfall of 177.17 mm and 177 rainy days, which looks incorrect for Morogoro, with irregular distribution (Gowing et al. 1999). The soil is a well-drained, acidic, kaolinitic clay with low essential nutrients (Kisetu and Mtakimwa 2020). Soil properties measured before transplanting are given in Table 1.

Table 1
Physical and chemical properties of the soil at the study site

Parameter	Unit	Value
Clay	%	54.12
Silt	%	9.64
Sand	%	36.24
Textural class	–	Clay
pH	–	5.2
Soil organic carbon	%	1.8
Total nitrogen	%	0.14
Available P	mg kg^{-1}	2.24
Cation exchange capacity	cmol kg^{-1}	7.28

2.2. Experimental design and treatments

The trial used a randomised complete block design (RCBD) with one factor, N rate, at four levels and four replications. Treatments were 0% (T0, control), 50% (T1), 100% (T2) and 150% (T3) of the recommended N rate, equal to 0, 50, 100 and 150 kg N ha^{-1} respectively (Table 2). Urea (46% N) was the N source. The recommended N rate corresponds to 6.2 g N $plant^{-1}$ at the 25 cm × 25 cm spacing used.Phosphorus (8.6 g P $plant^{-1}$) and potassium (2.6 g K $plant^{-1}$) were applied uniformly to all plots 5 days after transplanting, and N was applied 10 days after transplanting.

Table 2
Nitrogen treatments applied in the field experiment

Treatment	Proportion of recommended N	N rate (kg N ha^{-1})	N per plant (g)
T0	0%	0	0
T1	50%	50	3.1
T2	100%	100	6.2
T3	150%	150	9.3

Rice variety SARO 5 (TXD306), widely grown by Tanzanian smallholders, was used as the test crop. Seedlings were raised in a nursery for 23 days and then transplanted. Each plot measured 3.25 m × 1.75 m (5.69 m^2) with 1 m between plots, giving a trial area of 16 m × 8.5 m (136 m^2). Each plot held 40 plants at 25 cm × 25 cm spacing, and 12 plants per plot were selected for measurement using Python's

random module to avoid selection bias. Water and all nutrients other than N were managed uniformly across treatments, and plots were irrigated to at least field capacity.

2.3. Plant growth, leaf nitrogen and grain yield

Plant height (base to tip of the terminal leaf) and leaf length and width were measured with a tape measure on the 12 selected plants per plot at 40, 50, 60, 70, 80 and 90 days after planting (DAP). Leaf area was calculated as LA = L × W × k, where L is leaf length, W is leaf width and k is a shape coefficient of 0.75 (Aschonitis et al. 2014; Ma et al. 2023).

For leaf N, 20–30 fully developed top leaves were sampled at random from each plot at each of the six sampling dates. Samples were oven-dried at 70°C, ground, and analysed for total N by Kjeldahl digestion and distillation. Grain yield was determined at harvest.

2.4. Image acquisition

Images of rice leaves were taken with a GoPro Hero 10 camera (RGB) and a Mapir Survey3 camera (Red + Green + NIR) at the six sampling dates (40–90 DAP). To capture position-related variation, images were taken of the midsection, tip, canopy and vertical profile of the plants. All images were taken between 08:00 and 10:00 under natural daylight to limit variation in illumination. No radiometric calibration panel was used for the NDVI camera, and no lens-distortion correction was applied to the GoPro images (see Section 5).

2.5. Dataset preparation and quality control

Image labels correspond to the N treatment applied to the plot of origin (T0–T3) and images are stored in folders named accordingly. The raw collection contained 12,000 images (6,000 RGB and 6,000 NDVI). Corrupted and blurred images were removed automatically (PIL *Image.verify()* and a blur check), followed by manual inspection to remove duplicates and images with inadequate leaf coverage. Remaining images were cropped to the leaf area. The final dataset contains 8,000 JPG images (4,000 RGB and 4,000 NDVI) of varying pixel dimensions.

2.6. Statistical analysis

Data were organised in Microsoft Excel and analysed in R version 4.3.2 (R Core Team). For each sampling date, one-way ANOVA appropriate to the RCBD tested the effect of N rate on plant height, leaf area and leaf N content; grain yield was analysed in the same way. Means were separated with Tukey's honestly significant difference (HSD) test at $P \leq 0.05$. Figures were drawn with the ggplot2 package; error bars show ± standard error (n = 4).

2.7. Validation of the dataset: NDVI–tissue N relationship and baseline classifier

To check that the imagery carries agronomically meaningful information, mean NDVI was extracted from the Mapir Survey3 images and related to laboratory-measured leaf tissue N across all treatments and

sampling dates using Pearson correlation and simple linear regression. As a baseline for future users, a MobileNetV3 classifier (*mobilenetv3_large_100*) was trained to classify the four N classes, with performance estimated by 5-fold stratified cross-validation and reported as accuracy, precision, recall and F1-score.

3. Results

3.1. Agronomic response to nitrogen

Nitrogen rate significantly affected plant height, leaf area, leaf tissue N and grain yield (Table 3; Figs. 1–4). At 90 DAP, plant height was 81.13 cm in T0, 112.45 cm in T1, 123.05 cm in T2 and 128.72 cm in T3. Leaf area followed the same pattern (43.12, 66.22, 82.00 and 85.63 cm^2), as did leaf tissue N (1.33, 2.82, 4.88 and 5.09%) and grain yield (33.58, 48.52, 63.63 and 64.94 g $plot^{-1}$.

Table 3
Treatment means for plant height, leaf area and leaf tissue N at 90 DAP, and for grain yield at harvest. Letters indicating significant differences are given in Figs. 1–4

Treatment	Plant height (cm)	Leaf area (cm^2)	Leaf N (%)	Grain yield (g $plot^{-1}$)
T0 (0 kg N ha^{-1})	81.13	43.12	1.33	33.58
T1 (50 kg N ha^{-1})	112.45	66.22	2.82	48.52
T2 (100 kg N ha^{-1})	123.05	82.00	4.88	63.63
T3 (150 kg N ha^{-1})	128.72	85.63	5.09	64.94

Plant height differed significantly among treatments on every sampling date ($P < 0.05$). T0 plants were consistently the shortest, and differences among treatments were largest between 50 and 70 DAP, the phase of rapid vegetative growth. T2 and T3 reached similar maximum heights at most dates, so N above the recommended rate gave no further gain in height (Fig. 1).

Leaf area also differed significantly at every sampling date. T0 showed severely restricted leaf expansion, T1 was intermediate, and T2 and T3 reached the largest leaf area, particularly during 60–80 DAP, with no significant difference between them at most dates (Fig. 2). Adequately fertilised plants therefore developed a larger canopy and greater photosynthetic capacity.

Leaf tissue N differed significantly among treatments throughout the season (Fig. 3). Early in growth (40–50 DAP), T0 was critically deficient, while fertilised treatments had N status proportional to the rate applied. Tissue N peaked during mid-growth (60–70 DAP) and declined across all treatments toward maturity (80–90 DAP), which is consistent with remobilisation to reproductive organs and senescence, although treatment differences persisted.

Grain yield differed significantly among treatments (Fig. 4). T0 gave the lowest yield, T1 an intermediate yield, and T2 and T3 the highest yields, which were statistically similar.

3.2. Relationship between NDVI and leaf tissue nitrogen

Mean NDVI was strongly and positively correlated with laboratory-measured leaf tissue N across the four treatments and six sampling dates (Pearson r = 0.893, R^2 = 0.798, P < 0.001; Fig. 5a). Treatment means rose steadily from T0 (mean NDVI ≈ 0.21; tissue N ≈ 1.5%) to T3 (mean NDVI ≈ 0.63; tissue N ≈ 5.7%). The R^2 exceeds the value of 0.70 often cited as acceptable for spectral–agronomic relationships in precision agriculture (Khose and Mailapalli 2024), showing that the NDVI images carry quantitative information on N status.

3.3. Baseline classification

The baseline MobileNetV3 classifier, evaluated by 5-fold stratified cross-validation, reached an accuracy of 75.3% (precision 75.28%, recall 75.30%, F1-score 75.21%; Table 4, Fig. 5b). Classification was most accurate for T0 (86.4% of T0 samples correct), and errors occurred mainly between neighbouring N levels (Fig. 5b). This is consistent with the large spectral separation between T0 and the fertilised treatments shown in Section 3.2.

Table 4
Baseline MobileNetV3 classification performance (5-fold stratified cross-validation)

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
MobileNetV3 (mobilenetv3_large_100)	75.30	75.28	75.30	75.21

3.4. Description and availability of the dataset

The dataset is organised into four zipped folders, one per N treatment (0%, 50%, 100% and 150% of the recommended rate). Each folder contains separate RGB and NDVI sub-folders with 1,000 images each, giving 8,000 images in total (Table 5). Images are in JPG format and are named by class and a unique identifier. The dataset is deposited in the Zenodo repository and is openly available https://doi.org/10.5281/zenodo.17983378 (see Data availability).

Table 5
Distribution of the image dataset across nitrogen treatments and image types after quality control

Nitrogen treatment	RGB images	NDVI images	Total images
T0 (0 kg N ha^{-1})	1,000	1,000	2,000
T1 (50 kg N ha^{-1})	1,000	1,000	2,000
T2 (100 kg N ha^{-1})	1,000	1,000	2,000
T3 (150 kg N ha^{-1})	1,000	1,000	2,000
Total	4,000	4,000	8,000

4. Discussion

This study produced an open image dataset together with the agronomic and tissue N measurements needed to interpret it. Pairing image acquisition with ground-truth measurements allows visual and spectral symptoms of N deficiency to be tied to measured N status, which strengthens the dataset for machine-learning use (Cunha et al. 2021).

The control treatment showed stunted growth, reduced leaf expansion, chlorosis and premature senescence, as expected from the roles of N in chlorophyll synthesis, protein formation and enzyme production (Manni and Sharma 2023; Shrestha et al. 2022). The main agronomic finding was that the recommended rate (T2) and the 150% rate (T3) gave statistically similar plant height, leaf area, tissue N and grain yield at most sampling dates. In this single-site, single-season trial, N above the recommended rate therefore gave no measurable benefit, while the 50% rate (T1) clearly limited growth and yield. Applying the 150% rate would cost 50% more N fertiliser without a yield gain, and excess N carries a risk of leaching and nitrous oxide emissions (Liu et al. 2025). These economic and environmental implications are inferred from the agronomic results and were not tested directly. Because only one site, one variety and one season were studied, they should be confirmed across more environments before being generalised.

The strong NDVI–tissue N relationship (R^2 = 0.798) and the 75.3% baseline accuracy show that the images contain real signal on N status. The classifier's confusion between neighbouring N levels is expected, because adjacent treatments differ gradually and, with plot-level labels, some plants within a plot will not match the treatment label. Accurate, image-based N diagnosis built on this dataset could support affordable decisions on fertiliser use by smallholder farmers, extension officers and other stakeholders (Khose and Mailapalli 2024).

5. Limitations and recommendations for dataset users

Illumination and calibration. No radiometric calibration panel was used for NDVI capture, so NDVI values may shift under changing cloud cover. Users are advised to apply per-image histogram equalisation, relative NDVI scaling (normalising values within each image to the range 0–1) and channel-wise z-score normalisation computed over the whole dataset before training.

Lens distortion. No geometric lens-distortion correction was applied to the GoPro Hero 10 images before cropping. The wide-angle lens may affect the edges of some images, which may have contributed to the moderate accuracy of the baseline classifier.

Label noise. Labels are based on the treatment applied to each plot, not on verified plant-level health, so some images may not match their label. Users should consider noise-robust loss functions such as Symmetric Cross-Entropy or Generalised Cross-Entropy, or semi-supervised or active-learning approaches that identify and exclude anomalous samples within each class.

Scope. The images come from one site, one variety (SARO 5) and one season, and show disease-free leaves; models trained on them need testing on other varieties, environments and conditions.

6. Conclusions

We present an open dataset of 8,000 paired RGB and NDVI rice leaf images (1,000 per class and image type) from plants grown under four N rates, linked to measurements of plant height, leaf area, leaf tissue N and grain yield. In this trial, the recommended N rate (6.2 g N $plant^{-1}$) gave growth and yield similar to the 150% rate, whereas the 0% and 50% rates caused clear deficiency. Mean NDVI was strongly correlated with tissue N ($r = 0.893$), and a baseline MobileNetV3 classifier achieved 75.3% accuracy. The dataset provides a starting point for deep-learning tools that detect N deficiency in rice using smartphone cameras or low-cost imaging systems. Future work should extend the dataset to more sites, varieties and seasons, and add radiometric calibration and plant-level labels.

Abbreviations

ANOVA
analysis of variance
DAP
days after planting
HSD
honestly significant difference
K
potassium
N
nitrogen
NDVI

Normalised Difference Vegetation Index
NIR
near-infrared
P
phosphorus
RCBD
randomised complete block design
RGB
Red-Green-Blue.

Declarations

Ethics approval and consent to participate

Not applicable. This study did not involve human participants, human data or animals. The field experiment was conducted on a university research farm, and the rice variety SARO 5 (TXD306).The study complied with relevant institutional, national and international guidelines and legislation.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests. Confirm for all three authors

Funding

This research received no specific funding

Author Contribution

L.N.N. conceived the study together with K.G.F. and M.J.S., designed and conducted the field experiment, collected the plant and image data, prepared and quality-checked the dataset, performed the statistical analysis, and wrote the first draft of the manuscript. K.G.F. supervised the image acquisition and dataset preparation and the baseline classification analysis, and reviewed and edited the manuscript. M.J.S. supervised the agronomic and soil components of the experiment and reviewed and edited the manuscript. All authors read and approved the final manuscript.

Acknowledgement

First and foremost, I give all glory and honour to the Almighty God for His boundless mercy, guidance, and strength throughout my academic journey. His grace enabled me to persevere and complete this dissertation.I am deeply grateful to my brother Denis Gerson Ndiwaita for his unwavering love, guidance, and generous financial support that made my Master's studies possible.My heartfelt appreciation goes to my supervisors, Dr. Kadeghe Fue and Dr. Mawazo J Shitindi of the Department of Soil and Geological Sciences, for their expert guidance, constructive feedback, and tireless support throughout my research. Their mentorship was pivotal in shaping the direction and quality of this work.

Data Availability

i have generated imagery dataset of RGB and NDVI which are Publicly available at: https://doi.org/10.5281/zenodo.17983378

References

1. Akintuyi OB (2024) AI in agriculture: a comparative review of developments in the USA and Africa. Research Journal of Science and Engineering 10(2):60–70
2. Alfred R, Obit JH, Chin CP-Y et al (2021) Towards paddy rice smart farming: a review on big data, machine learning, and rice production tasks. IEEE Access 9:50358–50380. https://doi.org/10.1109/ACCESS.2021.3069449
3. Aschonitis VG, Papamichail DM, Lithourgidis A, Fano EA (2014) Estimation of leaf area index and foliage area index of rice using an indirect gravimetric method. Communications in Soil Science and Plant Analysis. https://doi.org/10.1080/00103624.2014.907917
4. Aulakh MS (1996) Nitrogen losses and fertiliser N use efficiency in irrigated porous soils. Nutrient Cycling in Agroecosystems 47:197–212
5. Aworka R et al (2022) Agricultural decision system based on advanced machine learning models for yield prediction: case of East African countries. Smart Agricultural Technology 2:100048. https://doi.org/10.1016/j.atech.2022.100048
6. Chianu JN, Chianu JN, Mairura F (2012) Mineral fertilizers in the farming systems of sub-Saharan Africa. A review. Agronomy for Sustainable Development 32:545–566
7. Christiaan Struik P et al (2022) Enhancing sorghum yield through efficient use of nitrogen – challenges and opportunities. Frontiers in Plant Science 13:845443. https://doi.org/10.3389/fpls.2022.845443
8. Coggins S, McDonald AJ, Silva JV et al (2025) Data-driven strategies to improve nitrogen use efficiency of rice farming in South Asia. Nature Sustainability 8(1):22–33. https://doi.org/10.1038/s41893-024-01496-3

9. Cunha J, Gaspar PD, Assunção E, Mesquita R (2021) Prediction of the vigor and health of peach tree orchard. Lecture Notes in Computer Science 12951:541–551. https://doi.org/10.1007/978-3-030-86970-0_38
10. Fue KG, Baitu GP, Jokonya O, Banwart S, Korsten L (2025) Digitalization of precision fertilization in East Africa: adoption, benefits and losses. Frontiers in Sustainable Food Systems 9. https://doi.org/10.3389/fsufs.2025.1497577
11. Gowing J, Mahoo H, Mzirai O, Hatibu N (1999) Participatory irrigation management and rice production in the semi-arid zone of Tanzania
12. Jiang R, Sanchez-Azofeifa A, Laakso K et al (2021) UAV-based partially sampling system for rapid NDVI mapping in the evaluation of rice nitrogen use efficiency. Journal of Cleaner Production 289:125705
13. Karunathilake EMBM, Le AT, Heo S, Chung YS, Mansoor S (2023) The path to smart farming: innovations and opportunities in precision agriculture. Agriculture 13(8):1593
14. Kavishe R (2021) Assessing adoption and water productivity of the system of rice intensification under farmer-led irrigation system in northern Tanzania. Thesis, Nelson Mandela African Institution of Science and Technology
15. Khose SB, Mailapalli DR (2024) UAV-based multispectral image analytics and machine learning for predicting crop nitrogen in rice. Geocarto International 39(1). https://doi.org/10.1080/10106049.2024.2373867
16. Kisetu E, Mtakimwa ZS (2020) Incorporating pigeon pea compost with Minjingu fertiliser brands to determine their effects on maize production in Morogoro, Tanzania
17. Kumar N, Chhokar RS, Meena RP et al (2021) Challenges and opportunities in productivity and sustainability of rice cultivation system: a critical review in Indian perspective. Cereal Research Communications 1–29
18. Laizer H, Mduma N, Machuve D, Maganga R (2024) Common beans imagery dataset for early detection of bean rust and bean anthracnose diseases. Data in Brief 54:110508. https://doi.org/10.1016/j.dib.2024.110508
19. Li S, Xu L, Jing Y, Yin H, Li X, Guan X (2021) High-quality vegetation index product generation: a review of NDVI time series reconstruction techniques. International Journal of Applied Earth Observation and Geoinformation 105:102640. https://doi.org/10.1016/j.jag.2021.102640
20. Liu Y, Chen Y, Peng D et al (2025) Determining the critical nitrogen application rate for maximizing yield while minimizing NO3−-N leaching and N2O emissions in maize growing on purple soil. Agronomy 15(6). https://doi.org/10.3390/agronomy15061358
21. Ly R (2021) Machine learning challenges and opportunities in the African agricultural sector – a general perspective. arXiv:2107.05101
22. Ma P, Zhang K, Liao X et al (2023) Effects of nitrogen application rate on dry matter weight and yield of direct-seeded rice under straw return

23. Manni TB, Sharma R (2023) Assessing the optimal level of NPK fertilisers for enhancing production of spring rice in Kanchanpur, Nepal. Big Data in Agriculture 5(1):36–42. https://doi.org/10.26480/bda.01.2023.36.42
24. Mduma N, Laizer H (2023) Machine learning imagery dataset for maize crop: a case of Tanzania. Data in Brief 48:109108. https://doi.org/10.1016/j.dib.2023.109108
25. Mgohele RN, Massawe BHJ, Shitindi MJ, Sanga HG, Omar MM (2025) Land suitability assessment for sisal production: a machine learning and analytical hierarchy process integrated approach. Soil Advances 3. https://doi.org/10.1016/j.soilad.2025.100048
26. Ministry of Agriculture (2021) Tanzania sample census of agriculture 2019/20. United Republic of Tanzania
27. Ndiwaita LN, Fue KG, Shitindi MJ (2025) Rice leaf RGB and NDVI imagery dataset under four nitrogen levels. Zenodo. https://doi.org/10.5281/zenodo.17983378
28. Nikitha S, Prabhanjan S, Sathyanarayan A (2025) Plant nutritional deficiency detection: a survey of predictive analytics approaches. Iran Journal of Computer Science 8(1):83–101. https://doi.org/10.1007/s42044-024-00209-5
29. Savci S (2012) An agricultural pollutant: chemical fertilizer. International Journal of Environmental Science and Development 3(1):73–80
30. Sekiya N, Asano A, Peter MA et al (2025) Effects of nitrogen application in upland rice cultivars: balancing sink-source relationships for sustainable yield in water-limited environments. Field Crops Research 332:110012. https://doi.org/10.1016/j.fcr.2025.110012
31. Sharma K, Shivandu SK (2024) Integrating artificial intelligence and Internet of Things (IoT) for enhanced crop monitoring and management in precision agriculture. Sensors International 5:100292. https://doi.org/10.1016/j.sintl.2024.100292
32. Sharma M, Nath K, Sharma RK, Kumar CJ, Chaudhary A (2022) Ensemble averaging of transfer learning models for identification of nutritional deficiency in rice plant. Electronics 11(1):148
33. Shrestha J, Karki TB, Hossain MA (2022) Application of nitrogenous fertiliser in rice production: a review. Journal of the Nepal Agricultural Research Council 8:16–26. https://doi.org/10.3126/jnarc.v8i.44815
34. Sithole A, Olorunfemi OD (2024) Sustainable agricultural practices in sub-Saharan Africa: a review of adoption trends, impacts, and challenges among smallholder farmers. Sustainability 16(22):9766. https://doi.org/10.3390/su16229766
35. Skiba U, Rees B (2015) Nitrous oxide, climate change and agriculture. CAB Reviews. https://doi.org/10.1079/PAVSNNR20149010
36. Sutton CL (2015) Impact of management on soil fertility and rice yields in smallholder farms in Tanzania. MSc thesis, The Ohio State University
37. Vlek PLG, Byrnes BH (1986) The efficacy and loss of fertilizer N in lowland rice. Fertilizer Research 9:131–147

38. Wanasundara JPD, Kumagai H, Kasch AP, Van Gels L, Peirce S (2024) Rice: a source of plant protein with many valuable processing co-products. In: Sustainable protein sources. Elsevier, pp 55–75
39. Zheng H, Cheng T, Li D et al (2018) Evaluation of RGB, color-infrared and multispectral images acquired from unmanned aerial systems for the estimation of nitrogen accumulation in rice. Remote Sensing 10(6):824

Figures

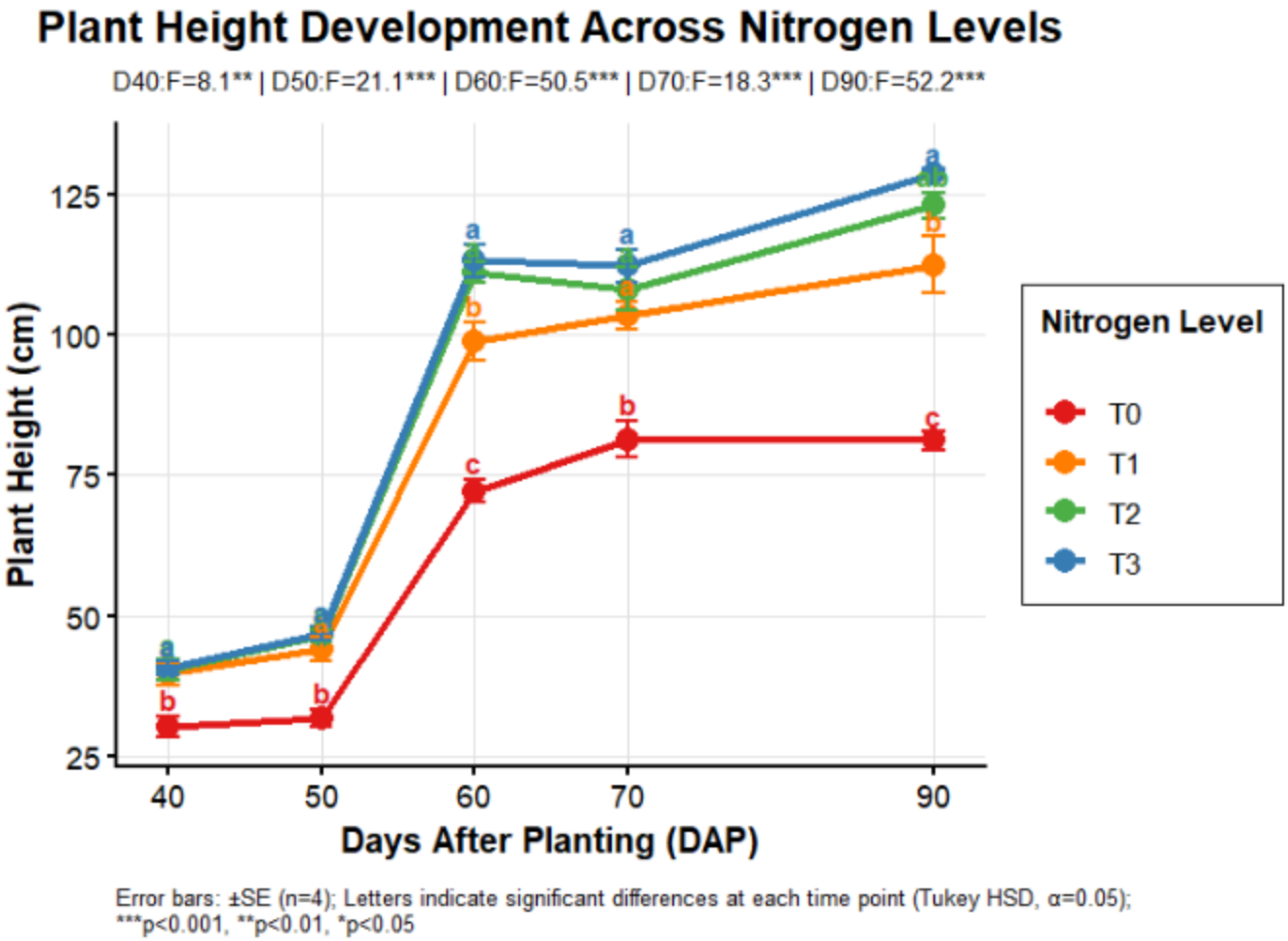


Figure 1

Plant height of rice under four nitrogen levels at 40, 50, 60, 70, 80 and 90 days after planting. Different letters at each time point indicate significant differences among treatments (Tukey's HSD, P ≤ 0.05)

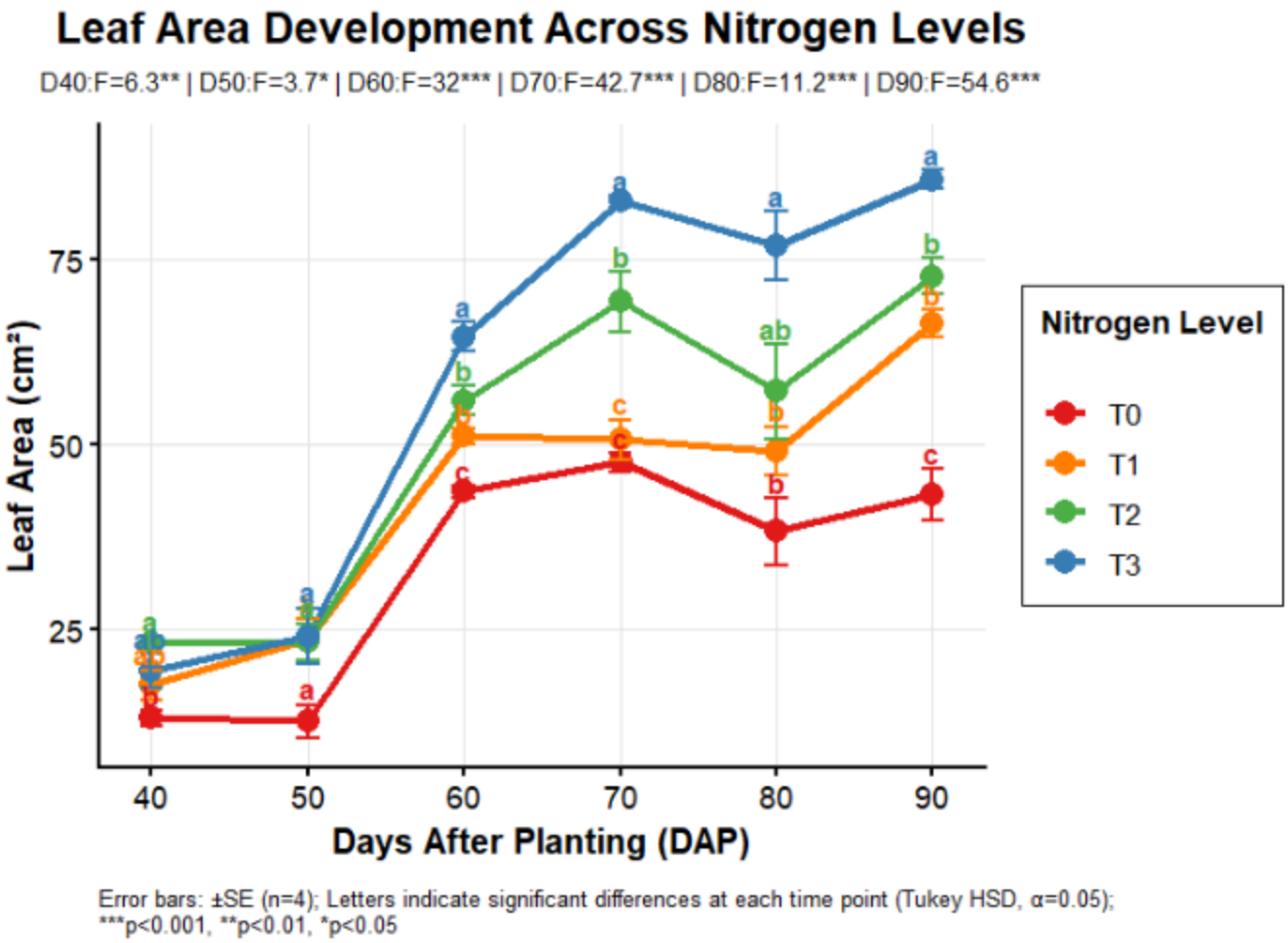


Figure 2

Leaf area of rice under four nitrogen levels at 40–90 days after planting. Different letters at each time point indicate significant differences among treatments

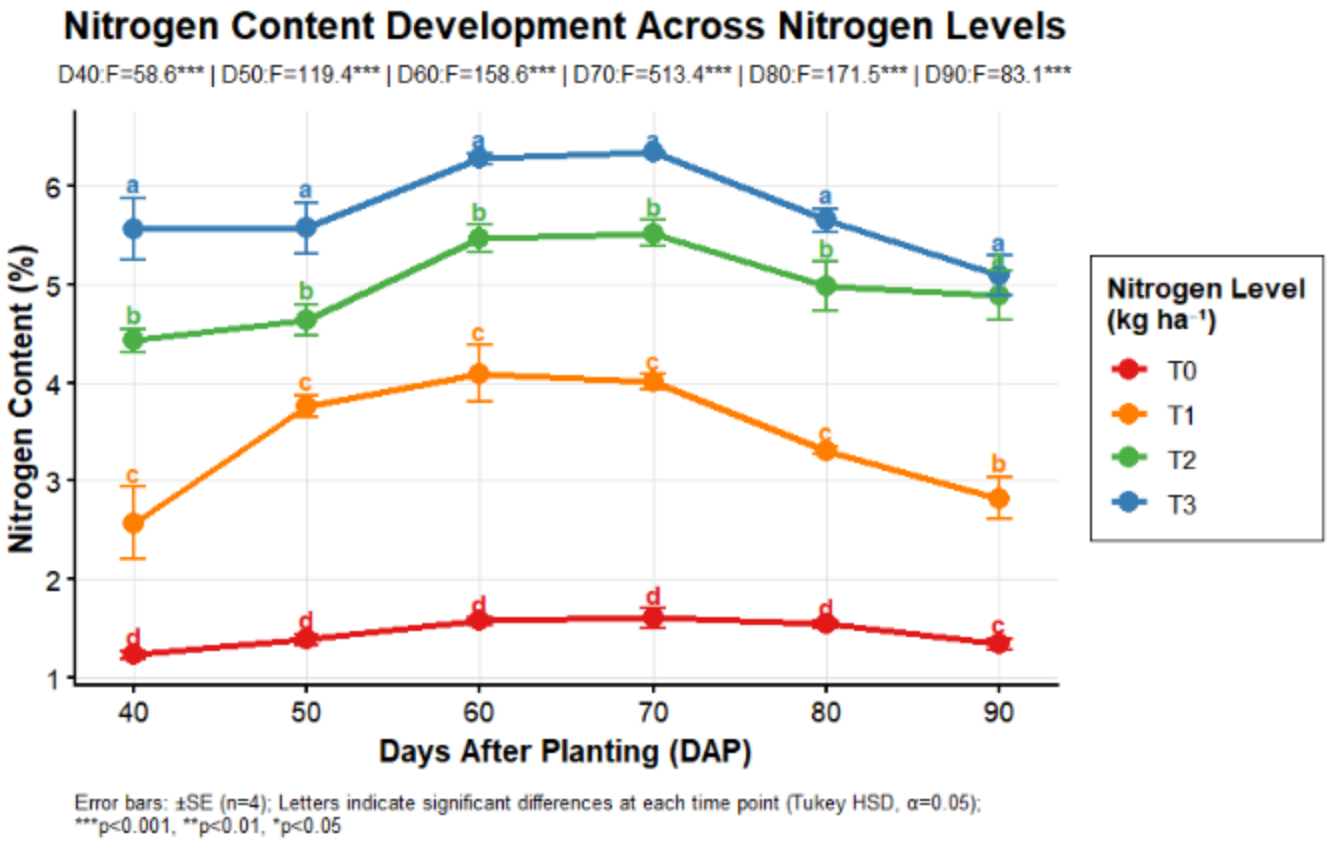


Figure 3

Leaf tissue nitrogen content of rice under four nitrogen levels at 40–90 days after planting. Different letters at each time point indicate significant differences among treatments

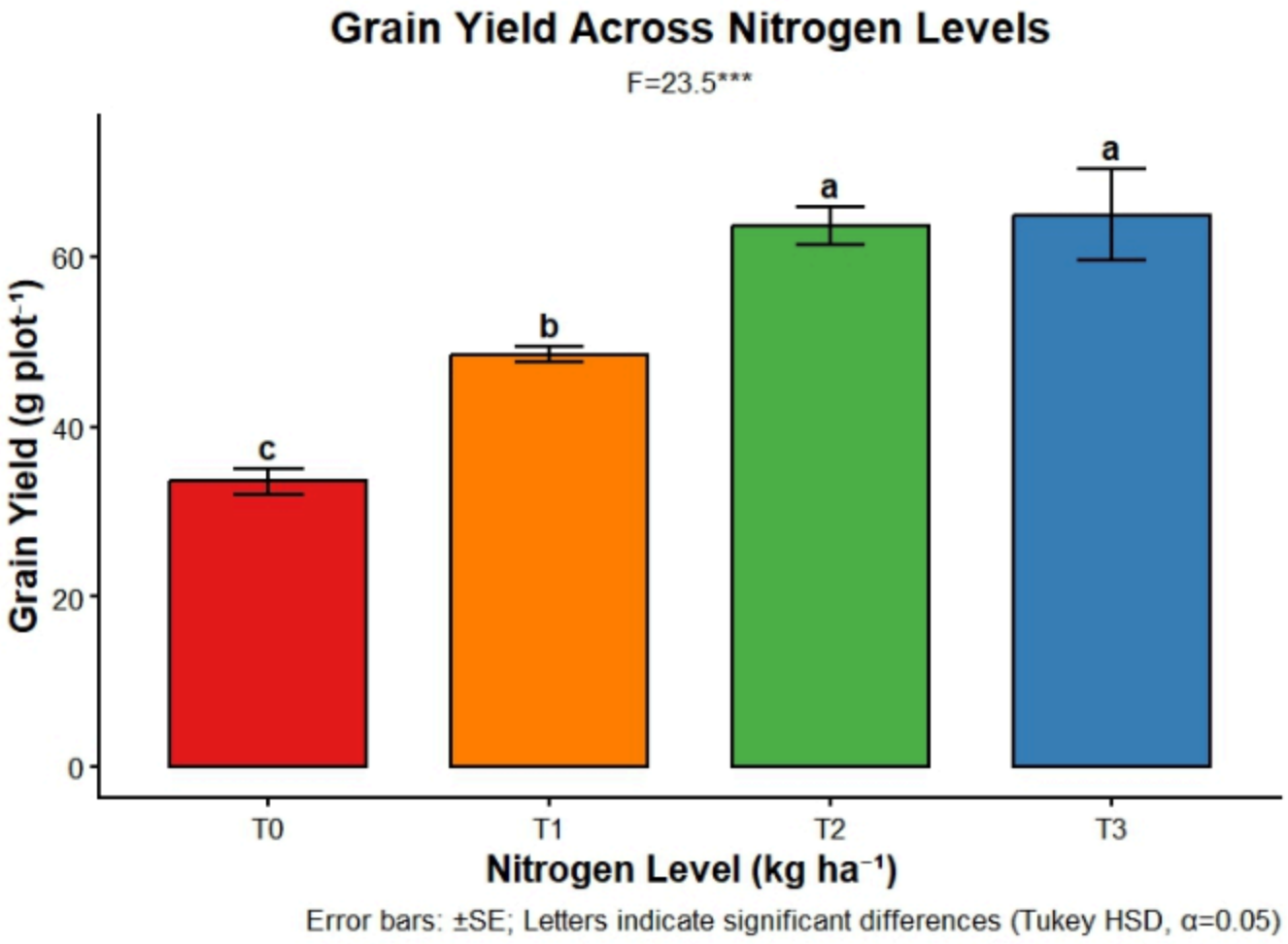


Figure 4

Grain yield of rice under four nitrogen levels. Different letters above bars indicate significant differences among treatments.

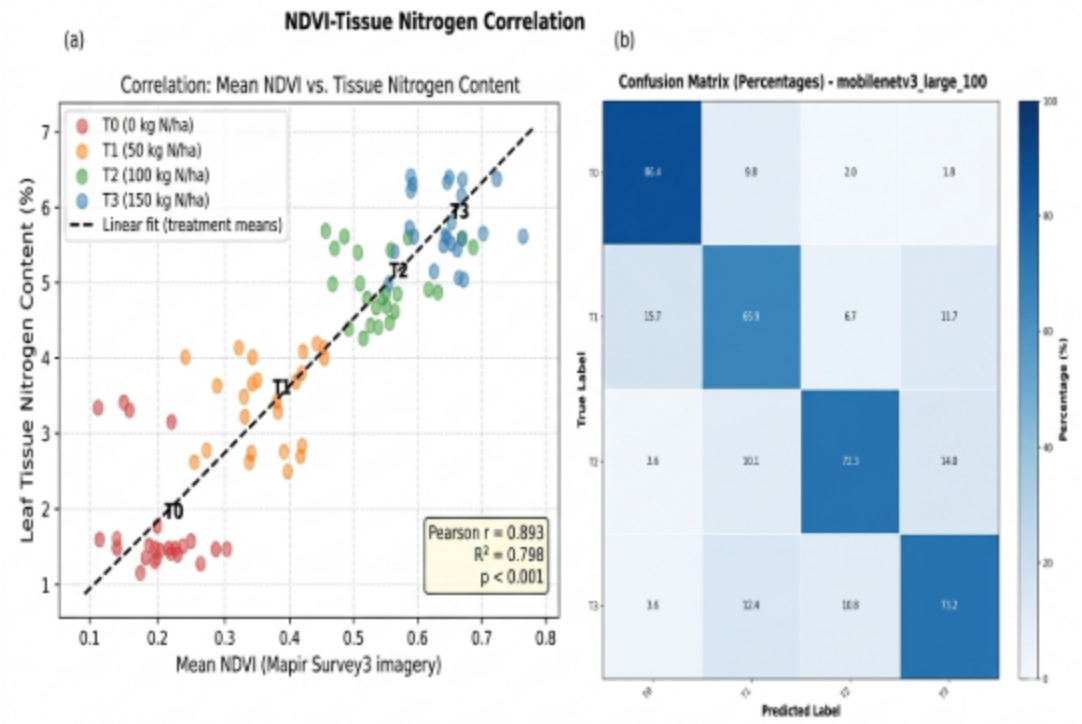


Figure 5

(a) Relationship between mean NDVI (Mapir Survey3 images) and leaf tissue nitrogen content across four nitrogen treatments and six growth stages (Pearson r = 0.893, R^2 = 0.798, P < 0.001); each point is one replicate measurement and treatment means are annotated. (b) Confusion matrix (%) of the baseline MobileNetV3 (mobilenetv3_large_100) classifier from 5-fold cross-validation (overall accuracy 75.3%)