Minor Revisions Response

We sincerely thank the reviewer for the careful evaluation of our manuscript and for the constructive comments that have helped improve its clarity, methodological transparency, and overall quality. We appreciate the positive assessment of the study and the recognition that the revised manuscript satisfactorily addressed the previous major concerns. All comments have been carefully considered, and the manuscript has been revised accordingly.

For ease of review, all modifications in the marked-up version of the manuscript are highlighted in **blue**. The responses to the reviewer comments below are also in **blue**.

1. My one substantive condition is that the authors address spatial autocorrelation, which remains uncontrolled and unmentioned: stratified random folds can place spatially adjacent subplots in both training and validation, and this is the most likely reason the cross-validation R2 values (around 0.95) exceed the holdout values (around 0.87). A spatial block cross-validation or a Moran's I diagnostic would be ideal, but at minimum an explicit caveat in the Limitations is required, together with a sentence noting that the cross-validation gain may partly reflect spatial proximity rather than added training data alone. There is no Moran's I, no spatial block CV, and no honest caveat. If anything, line 312 to 313 ("testing samples were distributed throughout the experimental field") describes the exact configuration that allows adjacent train and test subplots. Per your own decision rule, this keeps the paper at minor revisions and the caveat should be a required condition. The Limitations section (6.5, lines 635 to 657) covers crops, sites, seasons, layer parameterization, radiometric calibration, and scan angle, but says nothing about spatial autocorrelation, spatially adjacent subplots, or leakage. CV R2 exceeds holdout R2: partially / unchanged. The discrepancy is acknowledged and explained at lines 397 to 399, but only with the benign "more training data per fold" interpretation. That is the same sentence you saw last round. No cautioning sentence about possible spatial leakage was added, so in substance this is unchanged.

Response

The manuscript has been revised to explicitly acknowledge this limitation in the Model Evaluation section , Lines (331-340) and in the Discussion under limitation on Lines (657 - 663). We now clarify that the cross-validation folds were stratified by measurement date but were not spatially blocked, and that some spatial dependence may therefore exist between the training and validation subsets. We also note that the dataset incorporated substantial temporal, structural, and management variability through repeated measurement campaigns spanning the growing season, multiple wheat genotypes, and contrasting nitrogen fertilization treatments, although some influence of spatial autocorrelation could not be avoided within the relatively small experimental field.

In addition, we now clarify that the cross-validation results should be interpreted primarily as a comparative assessment of sensor feature combinations rather than as fully

independent estimates of model generalizability. We further recommend that future studies evaluate the proposed sensor metrics and modeling framework using independent sites, growing seasons, and spatially blocked validation strategies where feasible.

Lines (410-413), accordingly, the Results section now states that the observed difference may reflect both the increased training data available within each fold and the potential influence of spatial dependence.

2. I would also ask them to clarify the activation function (as written it reads as a purely linear network, which contradicts the ANN versus MLR result). Activation function contradicts the nonlinearity claim. Line 290 states "A linear (identity) activation function was applied." With one to two hidden layers (line 290), a purely identity activation collapses the network to linear regression, which directly contradicts the claims that the ANN captures nonlinear interactions (lines 82 to 83 and 374) and the result that the 5-layer GF ANN (R2 = 0.88) beats the same-input MLR (R2 around 0.75) in Figure 8.

Response

Lines (287-298): The manuscript has been revised to accurately describe the ANN architecture. We now state that a logistic activation function was applied to the hidden neurons, while a linear activation function was used only for the output neuron (linear.output = TRUE) to support continuous regression. This correction accurately reflects the implementation in the neuralnet package and clarifies that the network retains its nonlinear modeling capability through the hidden layer rather than reducing to a multiple linear regression model.

3. I would also ask them to clarify the per-fold "best of 20" selection rule.

Response

Lines (323 – 328) The manuscript now explains that, within each cross-validation fold, 20 networks were trained using identical architectures and training settings to account for stochastic variation arising from random weight initialization. Because ANN performance can vary between runs due to random initial weights, repeated initializations were used to improve the consistency of comparisons among sensor combinations. The network achieving the lowest validation RMSE within each fold was then selected. We also acknowledge that, because the validation subset was used to select the best-performing initialization, the reported cross-validation performance may be slightly optimistic and should therefore be interpreted primarily as a comparative assessment of sensor feature combinations rather than as an estimate of model generalizability.

4. TIR weather stability: partially fixed, indirect only. Still answered with hardware and correction parameters: emissivity, air temperature, humidity, blackbody, heated shutter (lines 158 to 161). Air temperature and humidity do enter the correction, and the LST versus CWSI comparison (lines 446 to 449) is an indirect argument that meteorological variation did not dominate. But there is still no direct statement confirming comparable acquisition conditions across TIR campaigns (clear sky, wind, time relative to solar noon, logged conditions).

Response

(Lines 160 -167): expanded the description of the multispectral and thermal image acquisition conditions and thermal correction workflow and clarified the acquisition conditions near solar noon under predominantly clear-sky conditions

5. Equation A3 mismatch and typos (lines 737 to 739): the equation still uses GF_i while the definition below still reads "p_i is the proportion...", and "n is the tolat nmber" is still there. Untouched. Heading 4.8 still reads "Mode Evaluation" (line 307). Heading 5.7 still reads "Feature sptial temporal alignment" (line 514).

Response

Corrected typographical and formatting errors throughout the manuscript (including "Mode Evaluation," "Feature spatial temporal alignment," and Appendix equation formatting).

6. Intensity wavelength is inconsistent. Line 186 says 905 nm, while the abstract (line 29), Section 4.4 (line 248), and the Conclusion (line 671) all say 903 nm; the sensor spec is 897 to 907 nm (line 150). Pick one value.

Response

Corrected the LiDAR wavelength to ensure consistency throughout the manuscript.

7. Figure caption color coding contradicts itself. Figure 7 caption (line 338) says training is light green and testing is dark green; Figure 9 caption (lines 385 to 386) says training is dark green and testing is light green. One is wrong, and it matters for reading the train versus test bars.

Response

Revised Figure 9 caption to refer to the correct color coding.

8. Missing from the list entirely: Virtue et al. 2021 (cited line 161), Melitha et al. 2025 (line 81), Wang et al. 2024 (lines 81 and 593; the list has only Wang 2017 and Wang 2021).

Year mismatches: Dawidowicz cited as 2024 (line 86) but listed as 2025; You cited as 2007 (line 257) but listed as 2017. Morgan is cited as 2025 (lines 301, 310) but listed as "(n.d.)". Spelling: the list entry "Saskai, T." should be "Sasaki" (cited as Sasaki, line 258).

Response

The bibliography has been thoroughly checked and updated to correct all inconsistencies identified by the reviewer. Missing references (e.g., Virtue et al., 2021 and Wang et al., 2024) have been added, publication years and author names have been corrected (e.g., Dawidowicz, You, Morgan, and Sasaki), and formatting has been standardized throughout the reference list. After revisiting its relevance, Melitha et al. (2025) was removed from the manuscript, as it was not considered essential to support the discussion.

9. Beyond the revisions requested by the reviewer, we made several minor improvements to the figures to enhance consistency and readability. These included refinements to the formatting and layout, as well as an improved presentation of the workflow figure (Figure 4). We also standardized terminology and acronym usage throughout the manuscript and figures (e.g., consistently using crop height (CH) rather than crop height model (CHM) where referring to the predictor variable). These modifications are editorial in nature and do not affect the results, interpretation, or conclusions of the study.