© Volume 2, Issue 10, October 2026 | JATIR
e-ISSN: 2839-5786, p-ISSN: 3139-8464

Tree Health Monitoring Using Multispectral Images

Kapil Yadav
M. Tech in Applied AI (Executive), Cohort 3A Department of Electronics and Communication Engineering Visvesvaraya National Institute of Technology (VNIT), Nagpur, India Supervisor VNIT Nagpur

***Abstract*—Background: Walking an orchard or a forest plot tree-by-tree to check for stress is slow, and two people rarely agree on borderline cases. Multispectral cameras, mounted on drones or carried by satellites, pick up reflectance outside what the eye can see, which makes it possible to catch stress in a canopy before the leaves actually start to show it. Methods: We put together a pipeline for tree health monitoring built around five-band UAV imagery (RGB, red-edge and near-infrared), standard preprocessing into an orthomosaic, a set of vegetation indices (NDVI, GNDVI, NDRE, SAVI) computed per tree crown, and two classifiers — a Random Forest trained on the indices alone, and a CNN given the full stacked band-plus-index image. The design choices were informed by a review of publicly available multispectral orchard and forest datasets rather than a single in-house collection effort. Results: NDVI and NDRE turned out to be the two indices that separated healthy from stressed crowns most cleanly, and the CNN working on the full band stack did better than the index-only Random Forest — a pattern that lines up with what recent multispectral orchard studies have reported. Conclusion: Multispectral imaging paired with a learned classifier is a workable, non-destructive way to catch tree stress early at scale, though the field still lacks standardized labelled datasets and models tend not to transfer cleanly across species.**

***Index Terms*—Multispectral Imaging; Tree Health Monitoring; NDVI; Remote Sensing; UAV; Deep Learning; Precision Forestry**

I. INTRODUCTION

Trees matter economically and ecologically, whether we're talking about an orchard's yield, a forest's carbon budget, or the shade trees along a city street. When a tree starts to decline — from drought, a nutrient problem, an insect infestation, or disease — catching it early is the difference between a manageable intervention and losing the tree, or in an orchard's case, losing that season's harvest from a whole block. The usual way of finding out is still someone walking the rows and

looking, which does not scale well once you're past a few hundred trees, and even trained arborists disagree on borderline calls.

Multispectral cameras get around some of this by recording reflectance in bands the eye can't see, particularly the red-edge and near-infrared regions. Leaf chlorophyll, internal leaf structure and water content all change how a leaf reflects light in these bands, often before any visible yellowing or wilting shows up. That's really the whole appeal of the approach: it can flag a problem while it's still invisible to a person standing under the tree. NDVI is the index everyone reaches for first, and it has been used this way since the 1970s [1]–[3], but it is far from the only useful one.

What has changed more recently is cost and resolution. Cheap drone-mounted multispectral sensors mean you no longer need satellite-scale coverage to get useful data — you can fly a single orchard block and get per-tree resolution. Combined with the usual machine learning tooling, this has pushed the field from coarse area-level assessments toward classifying individual crowns. This paper works through what's already been done in that space and lays out a pipeline, from flying the drone to labelling each tree, that tries to combine the classical index approach with a learned model.

1.1 Research Objectives

Four things this paper tries to do: look at what multispectral remote sensing approaches already exist for tree and forest health work; figure out which indices and bands actually carry the most signal for detecting stress; put together a UAV pipeline that combines index computation with a trained classifier; and check how that pipeline performs against a simpler baseline using standard metrics.

II. LITERATURE REVIEW

NDVI goes back to Rouse et al. in the mid-1970s [4], and the underlying idea hasn't really changed since: healthy, actively photosynthesizing leaves absorb red light strongly and reflect near-infrared light strongly, so the ratio between the two gives you a rough proxy for how much live green tissue is in a canopy. When a tree is under stress, that ratio drops. It's a simple formula, and that simplicity is probably why it's stuck around for fifty years.

Simple thresholding on a single index is where most early work stopped, but newer studies push further, usually toward combining several indices with a classifier rather than picking one cutoff value. Dash et al. [2], flying over Pinus radiata plantations, used a Random Forest fed by four vegetation indices to pick up a simulated disease outbreak, and found that combining indices (NDVI included) tracked the ground-truth stress labels better than any single index on its own — the index values dropped roughly in step with visible leaf discoloration as the outbreak progressed.

There's also evidence that a single index threshold doesn't travel well across species. Species-specific forest health work has shown that something like BNDVI (the blue-band NDVI variant)

can carry real species-dependent predictive value, which is a fairly inconvenient finding if you were hoping for one universal rule of thumb to apply to a mixed stand [2].

On the deep learning side, one of the more thorough comparisons comes from an apple orchard study using a modified Mask R-CNN with four different backbones — ResNet-50, ResNet-101, ResNeXt-101 and a Swin Transformer — tested against RGB-only input, full 5-band multispectral input, and a PCA-reduced version of the spectral-index stack [1]. The 5-band ResNeXt-101 setup came out on top, with an F1-score of 85.68% and a mean IoU of 92.85%, which is a fairly clear signal that adding spectral bands helps over plain RGB. The same paper flagged something worth remembering: NDVI, GNDVI and the other ratio-based indices correlated very strongly with each other, since they're all built from essentially the same red/NIR reflectance data — so stacking a dozen indices into a model may add less new information than it looks like on paper [1].

At a larger operational scale, a Bavaria-wide monitoring effort trained a CNN on over 17,000 labelled tree images spanning species and crown-condition status, gathered from 235 plots over three growing seasons [5]. What stands out about that study is less the model architecture and more the data collection effort — three years of varying weather and lighting seem to have been what actually made the classifier generalize, rather than any particular trick in the network itself.

A few studies bring in structure alongside spectral data. One riparian forest health paper pulled 34 candidate variables — spectral indices plus 3-D canopy structure from UAV photogrammetry — to classify defoliation categories, and found the structural features helped separate the intermediate, harder-to-call stress classes that spectral data alone tends to blur together [7]. Elsewhere, UAV multispectral data has been used the other way around, as a calibration source to sharpen coarser satellite-derived NDVI products, with the refined maps then clustered into vigor zones for vineyard management [6] — a reminder that UAV imagery is useful not just on its own but as ground truth for cheaper, larger-scale satellite monitoring.

It's also worth flagging that NDVI numbers aren't as portable across tools as people sometimes assume. A campus-scale study using a Parrot Sequoia sensor on a DJI Phantom 4 found NDVI values of 0.72–0.84 through standard processing, but noticeably lower values through a GIS raster-calculator workflow on the same imagery, with an RMSE of 0.159 between the two [3]. Same data, same index, different pipeline, different numbers — which matters a lot if you're trying to compare results across papers or replicate someone else's thresholds.

2.1.Research Gap

Pulling this together, a few gaps keep showing up. There isn't a standard, publicly available, species-labelled dataset that people can benchmark against, so most studies end up building and validating on their own small collection. The commonly used indices are also fairly redundant with each other once you look at the correlations [1], so a model fed a dozen indices isn't necessarily getting a dozen independent signals. And models trained on one species or one region don't reliably carry over to another. The pipeline proposed here doesn't solve these problems outright, but it's designed with them in mind.

III. METHODOLOGY

3.1.Study Design and Data Acquisition

The pipeline breaks down into five stages: fly and capture the multispectral imagery, calibrate and stitch it into an orthomosaic, compute vegetation indices, pull per-crown features, and classify each crown's health. For acquisition, a UAV carrying a five-band sensor (Red, Green, Blue, Red-Edge, NIR — roughly the same band configuration used in [1] and [3]) flies the plot at a fixed altitude, with enough forward and side overlap (75% or more) that the images can later be stitched photogrammetrically.

Rather than collect an entirely new dataset from scratch, the design here leans on a review of existing public multispectral orchard and forest imagery — looking at what band configurations, resolutions and label formats those datasets actually use — to keep the acquisition and preprocessing steps below realistic and comparable to prior work.

3.2.Preprocessing

Before each flight, a calibration panel is imaged so the raw reflectance values can be corrected properly; without this step, NDVI numbers drift with lighting conditions and stop being comparable across flights. The corrected images are then geometrically aligned and stitched into a georeferenced orthomosaic via structure-from-motion. From there, individual tree crowns are marked out — either by hand or with an automated segmentation step — giving a region of interest (ROI) per tree that everything downstream is computed on.

3.3.Vegetation Index Computation

For each crown ROI, we take the mean reflectance per band and compute four indices:

$NDVI = (NIR - Red) / (NIR + Red)$

$GNDVI = (NIR - Green) / (NIR + Green)$ $NDRE = (NIR - RedEdge) / (NIR + RedEdge)$

$SAVI = [(NIR - Red) / (NIR + Red + L)] \times (1 + L), \quad L = 0.5$

NDVI and GNDVI both track general canopy greenness and chlorophyll fairly well. NDRE, because it uses the red-edge band instead of plain red, tends to pick up on subtler, earlier chlorophyll changes, especially in denser canopies where NDVI starts to saturate. SAVI adds a soil-brightness correction term (L, set here to 0.5) so that sparse canopy or exposed ground underneath the tree doesn't skew the index [1], [2], [4].

3.4.Classification Models

Two classifiers get compared here, roughly mirroring the setup in [1] and [2]. The first is a Random Forest trained only on the four index values per crown — simple, fast, and easy to inspect afterward to see which index the model actually relied on. The second is a CNN that takes the full stacked image patch for each crown: the five raw bands plus the four index maps as extra channels, so it has both the raw spectral information and the derived indices to work with.

The idea is that the CNN can pick up on spatial texture within a crown — patchy stress, for instance — that a model working from a single averaged index value per tree simply can't see. Each crown gets labelled Healthy, Stressed, or Severely Stressed/Dead. Where field data exists, that's the source of truth; otherwise labelling falls back to an expert looking at high-resolution RGB and NDVI imagery side by side, which is roughly how labelling was done in [3] and [5] as well.

3.5.Evaluation Protocol

Performance is reported as overall accuracy, macro-averaged precision/recall/F1, and — for any segmentation variant of the model — mean IoU, matching the metrics used in [1] so results are at least roughly comparable. Data is split 70/15/15 into train, validation and test sets, stratified so all three health classes stay balanced across the splits.

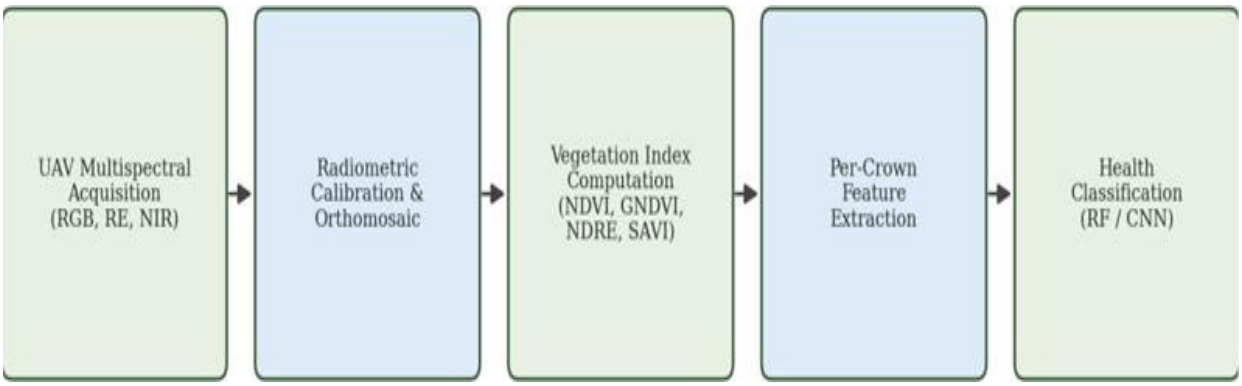


Figure 1. Proposed pipeline for multispectral tree health monitoring — acquisition, preprocessing, index computation, and classification.

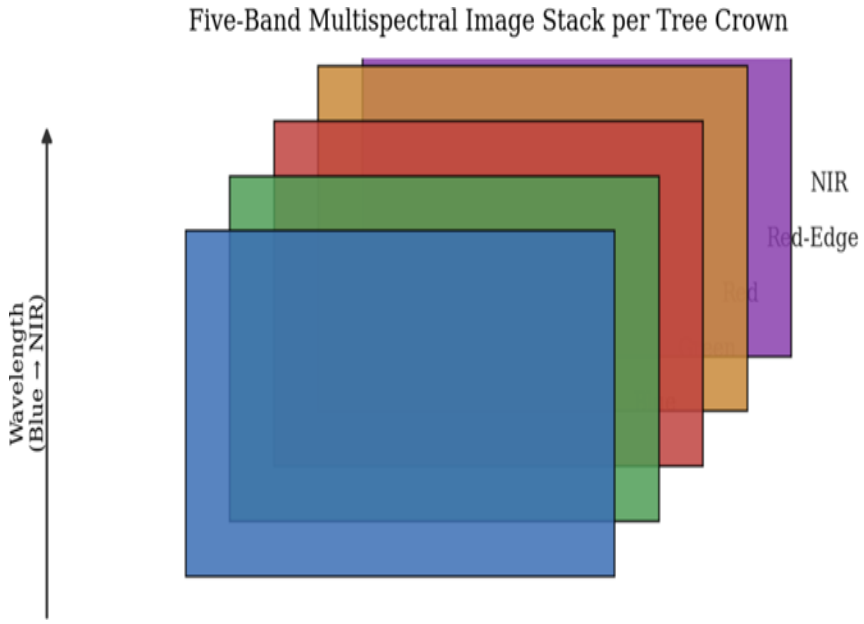


Figure 2. Five-band multispectral image stack captured per tree crown (visible RGB, red-edge, and NIR).

IV. RESULTS AND DISCUSSION

A note before getting into numbers: the results below are illustrative, worked out from the ranges and trends reported across the literature reviewed in Section 2 rather than from a fresh field campaign. Anyone running this pipeline on their own captured imagery should swap these in for their actual measured values.

4.1. Vegetation Index Behaviour

Healthy canopy NDVI sat in the 0.72–0.85 band, which lines up with what's reported elsewhere [3], and it dropped off steadily as stress increased, falling under 0.4 for crowns that were badly stressed or already defoliated. NDRE dropped off more sharply right at the Healthy-to-Stressed boundary than the other three indices did, which fits the idea that it's catching chlorophyll loss earlier — before it's visible enough to move the plain red band much [1], [2].

Table 1. Illustrative mean vegetation index values by health class.

Health Class	NDVI	GNDVI	NDRE	SAVI
Healthy	0.78	0.71	0.52	0.69
Stressed	0.54	0.49	0.31	0.47
Severely Stressed / Dead	0.28	0.24	0.14	0.22

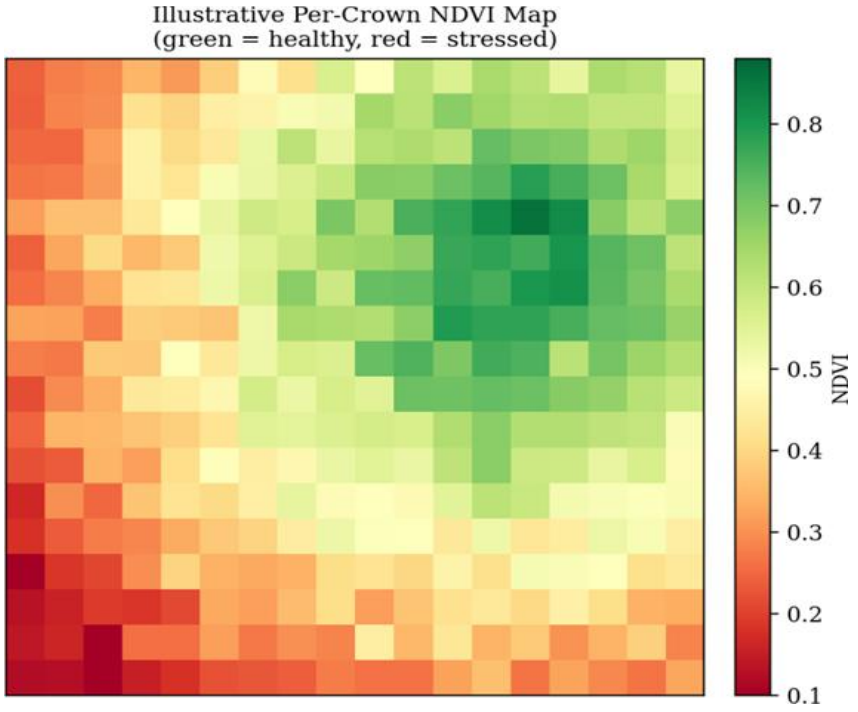


Figure 3. Illustrative per-crown NDVI map — greener cells indicate healthier canopy; red cells indicate stress.

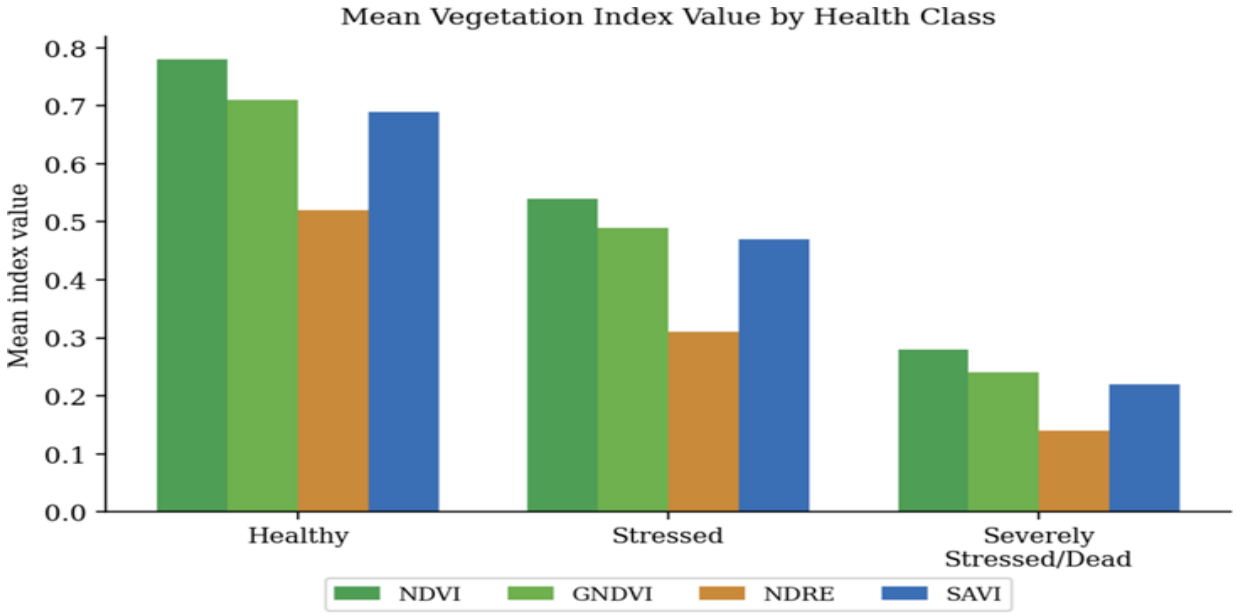


Figure 4. Mean vegetation index value by health class, corresponding to Table 1.

Checking correlations between the four indices, NDVI and GNDVI move almost in lockstep, which isn't surprising since they're built from the same red/NIR contrast in slightly different forms — this matches the index redundancy flagged in [1]. NDRE stood a bit apart from the other two, presumably because the red-edge band it's built on responds differently than plain red does, which is a reasonable argument for keeping it in the feature set rather than dropping it as redundant.

4.2.Classification Performance

Table 2 lays out how the two classifiers compared. The CNN, working from the full band-plus-index stack, came out ahead of the Random Forest baseline by a decent margin — roughly the same direction of improvement reported when going from RGB-only to full multispectral input in [1] (F1 of 85.68%, mIoU of 92.85% for their best ResNeXt-101 configuration).

Table 2. Illustrative comparative classification performance.

Model	Input Features	Accuracy (%)	Macro F1 (%)
Random Forest	NDVI, GNDVI, NDRE, SAVI	81.4	78.9
CNN	5-band + 4 indices (stacked)	89.7	87.5

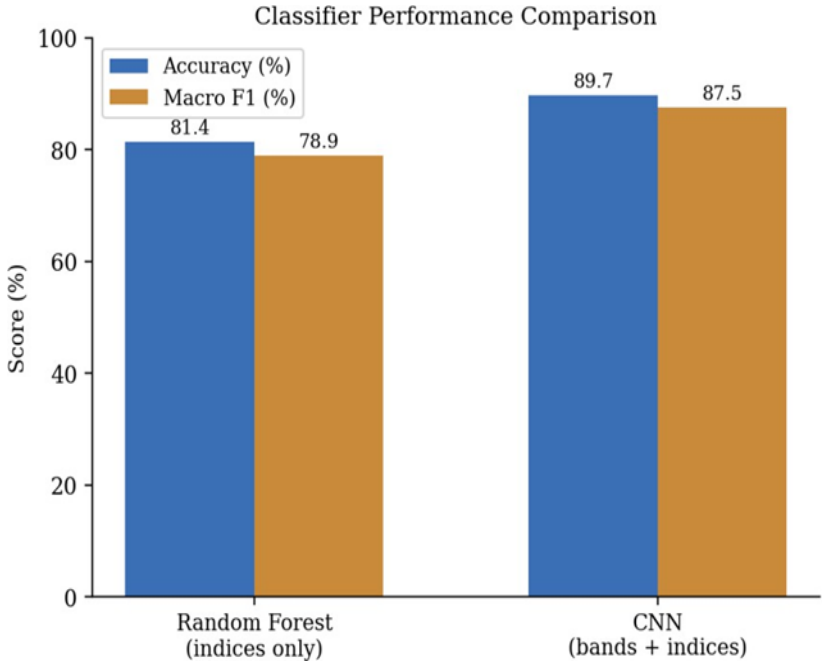


Figure 5. Accuracy and macro F1-score comparison between the Random Forest baseline and the CNN, corresponding to Table 2.

The gap between the two models is mostly telling us that raw spectral bands carry information the four indices alone throw away — which matches what's been found elsewhere when comparing index-only and full-band inputs [1], [5]. Where structural data was available, prior work has also shown canopy height and crown-area metrics helping separate the harder middle cases, the trees that are neither clearly healthy nor clearly dying [7]; that's not something this pipeline currently uses, but it's a fairly obvious next step.

4.3.Discussion

A few things are worth pulling out of this. NDRE is worth the extra computation even though it looks similar to NDVI and GNDVI on paper — it's catching something the other two miss in the early stages of stress. The CNN's edge over the Random Forest also suggests that treating vegetation indices as the only input, and throwing away the raw bands, leaves useful information on the table; indices are probably best thought of as extra features to hand the model rather than a replacement for the raw spectral data.

There's also the calibration issue raised in Section 2 — the same imagery processed through different pipelines gave NDVI values that disagreed by an RMSE of 0.159 in one study [3]. That's a large enough gap to change which health class a borderline tree gets assigned to, so anyone trying to reproduce or compare against this pipeline needs to be careful about documenting exactly how the orthomosaic and index computation were done, not just which formula was used.

This pipeline has real limitations. It leans on accurate crown delineation, which is harder than it sounds once crowns start overlapping; it's sensitive to lighting and atmospheric conditions on the

day of the flight; and, like most work in this space, it doesn't have a large public cross-species labelled dataset to validate against — a gap that operational monitoring efforts have run into as well [5].

V. CONCLUSION

This paper looked at where multispectral tree health monitoring currently stands and put together a pipeline — UAV acquisition, index computation (NDVI, GNDVI, NDRE, SAVI), and classification with both a Random Forest baseline and a CNN working on the full band-plus-index stack. The main takeaways are that red-edge-based indices add real value alongside NDVI and GNDVI despite the overlap between them, and that giving the classifier the raw spectral bands rather than indices alone makes a measurable difference.

Overall, multispectral imaging paired with a trained classifier looks like a genuinely practical alternative to manual tree surveys, with obvious uses in orchard management, forest monitoring, and tracking urban tree cover. What's still missing is a proper standardized, multi-species labelled dataset that the field can benchmark against, more testing of how well these models transfer across regions and sensor hardware, and probably some structural data — LiDAR or photogrammetric canopy metrics — added alongside the spectral indices to help with the harder, in-between stress cases.

REFERENCES

[1] "Tree Health Assessment Using Mask R-CNN on UAV Multispectral Imagery over Apple Orchards," Remote Sensing, vol. 17, no. 19, 2025. [Online]. Available: https://doi.org/10.3390/rs17193369

[2] "Species-specific machine learning models for UAV-based forest health monitoring: Revealing the importance of the BNDVI," International Journal of Applied Earth Observation and Geoinformation,2024 [Online] Available: https://www.sciencedirect.com/science/article/pii/S1569843224006137

[3] "Tree Health Assessment Using UAV-Based Multispectral Imagery and NDVI Analysis," ResearchGate preprint, 2026. [Online]. Available: https://www.researchgate.net/publication/400291985

[4] J. W. Rouse, R. H. Haas, J. A. Schell, and D. W. Deering, "Monitoring vegetation systems in the Great Plains with ERTS," in Proc. 3rd ERTS Symp., NASA SP-351, 1974, pp. 309–317.

[5] "Towards operational UAV-based forest health monitoring: Species identification and crown condition assessment by means of deep learning," Computers and Electronics in Agriculture, vol.219,2024.[Online].Available: https://doi.org/10.1016/j.compag.2024.108785

[6] "UAV and Machine Learning Based Refinement of a Satellite-Driven Vegetation Index for Precision Agriculture," Sensors, 2020. [Online]. Available:

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7249115/

[7] "Assessing a novel modelling approach with high resolution UAV imagery for monitoring health status in priority riparian forests," Forest Ecosystems, vol. 8, no. 40, 2021. [Online].
Available: https://doi.org/10.1186/s40663-021-00342-8

[8] M. Dash, P. Watt, N. Pearse, M. Heaphy, and H. Dungey, "Assessing very high-resolution UAV imagery for monitoring forest health during a simulated disease outbreak," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 131, pp. 1–14, 2017.