RGB–NDVI Channel Fusion for Citrus Leaf Disease Classification Using a Four-Channel ResNet50 and Sentinel-2 Satellite Data

Sanket R. Dhumal[1], Dr. Anil R. Karwankar[2] ,Vijay G. Nemane[3]
[1,2,3] *Department of Electronics and Telecommunication Engineering, Government College of Engineering,Chhatrapati Sambhajinagar, Maharashtra, India*

***Abstract*—Fungal and bacterial diseases and nutrient deficiency reduce citrus production in India by 30–40% annually. Early and precise identification is crucial for sustainable crop management, but traditional visual diagnosis is subjective and labour-intensive at scale. This study proposes a channel-level early fusion approach combining RGB leaf images with a spatially resolved Normalized Difference Vegetation Index (NDVI) map derived from Sentinel-2 Level-2A satellite data as a fourth input channel to a modified ResNet50 convolutional neural network. The 224×224 NDVI map was computed from Sentinel-2 tile T44QKJ (2023-02-12) covering the Nagpur citrus belt, providing region-specific spectral vegetation context. The model was evaluated on a publicly available citrus leaf benchmark dataset (38,432 annotated images, five classes), with both the RGB-only baseline and the proposed RGB+NDVI fusion model trained and tested across four random seeds to assess stability. The fusion model achieved 100.00 ± 0.00% weighted test accuracy across all seeds, compared to 99.99 ± 0.02% for the baseline. In the one seed where the baseline erred, the fusion model corrected all errors, and in no seed did the fusion model underperform; however, McNemar's exact test did not reach statistical significance (p = 0.25), reflecting near-ceiling performance on this benchmark. These results indicate that the satellite-derived NDVI channel provides complementary spectral context that never degrades classification, and that the proposed open-data pipeline is reproducible and suitable for low-resource precision agriculture deployment.**

***Index Terms*—Channel fusion, Citrus disease detection, Convolutional neural network, Deep learning, NDVI, Precision agriculture, Remote sensing, ResNet50, Sentinel-2.**

I. INTRODUCTION

Agriculture is the backbone of India, providing employment to more than 40% of the population and contributing approximately 18% of national GDP. Citrus fruits, especially Nagpur Mandarin (Citrus reticulata), are a major economic crop in the Vidarbha region of Maharashtra. Citrus orchards are, however, constantly exposed to foliar diseases such as Citrus Canker (Xanthomonas axonopodis), micronutrient deficiencies (zinc and iron), and multi-disease co-infections, which reduce productivity by 30–40% when uncontrolled [1,2].

Traditional disease identification relies on visual inspection by trained agronomists, which is subjective, slow, and difficult to scale across large orchards. The rise of precision agriculture using deep learning has opened up new opportunities for automated, real-time plant health monitoring. A number of studies have shown that RGB images of leaves are very useful for disease classification using convolutional neural networks (CNN) [1–4]. However, surface morphology can only be seen with RGB images; satellite imagery can reveal physiological changes that are not visible to the naked eye or traditional cameras [5].

The Sentinel-2 satellite mission's multispectral data offers atmospherically corrected surface reflectance at 10 m spatial resolution in 13 spectral bands. The Normalized Difference Vegetation Index (NDVI), calculated from the red and near-infrared bands, is a well-established proxy for photosynthetic activity and vegetation health [6,7]. Previous studies have shown that crop-level monitoring is valuable [8–10], but no prior study has used a spatially resolved NDVI spatial map derived from satellite observations as a dedicated fourth input channel to a ResNet50 classifier for leaf-

level citrus disease classification on a publicly available benchmark dataset. To fill this gap, this work proposes a channel-level early fusion architecture that introduces an NDVI spatial map, created by the authors from an actual Sentinel-2 Level-2A tile of the citrus belt in Nagpur, as an additional input channel to the RGB leaf image. The ResNet50 backbone is modified to accept four-channel input, with the fourth-channel weights initialised as the mean of the pretrained ImageNet RGB weights — a principled initialisation strategy that preserves transfer learning benefits while accommodating the additional spectral channel. This work makes the following contributions:
(1) A channel-level early fusion architecture that integrates a spatially resolved Sentinel-2 NDVI map as a fourth input channel to ResNet50 for citrus leaf disease classification, with principled fourth-channel weight initialisation.
(2) A fully reproducible pipeline built on two publicly available Kaggle datasets [11,12] and open-source tools (Python, PyTorch, rasterio), requiring no proprietary hardware or closed data.
(3) Multi-seed experimental validation (four independent random seeds) showing that the four-channel fusion model never underperforms the RGB-only baseline, and fully corrects baseline errors when they occur — including, in one seed, recovering three misclassifications between Citrus Canker and Nutrient Deficiency, two visually similar but agriculturally distinct conditions requiring different interventions.
(4) A comprehensive discussion of the design, limitations, and deployment pathway of the proposed framework for Indian precision agriculture.

II. LITERATURE REVIEW

The use of deep learning for automated detection of plant diseases is a well-established approach. Onker and Pandey [1] compared four architectures — VGG16, ResNet50, InceptionV3, and DenseNet — achieving 96–98% accuracy on the PlantVillage benchmark across 38 classes. Chow et al. [2] evaluated lightweight models such as MobileNetV2 and EfficientNet-B0 for edge deployment, achieving 95–99% accuracy with reduced computational cost. These studies demonstrate that RGB-based CNN classification is well established and does not require spectral data outside the visible spectrum. For crop yield prediction, Rohini et al. [3] proposed a hybrid CNN achieving 97.3% accuracy on multispectral leaf data using VGG16, VGG19, ResNet50 and DenseNet121. To classify plant health, Balasundaram et al. [5] proposed a pipeline combining NDVI estimation with Grounded-DINO and the Segment Anything Model (SAM), achieving 97.2% accuracy on a segmented plant leaf domain, showing that spectral features derived from NDVI contribute measurably to plant health class separability. Luo et al. [8] showed that spatial feature streams fused with spectral indices (NDVI, EVI) improve cropland segmentation by 3–5% compared to single-stream models using Sentinel-2 data. This fusion strategy is further supported by Li et al. [7], who showed that Sentinel-2 NDVI data captures macro-level vegetation trends that complement high-resolution field imagery. Deka and Chakraborty [13] proposed channel-level fusion for litchi fruit detection using multimodal UAV sensors and found that channel-level fusion outperforms single-modality baselines. Shafi et al. [10] showed that multi-modal crop health mapping (remote sensing, IoT sensors and machine learning) consistently outperforms single-modality approaches. Notably, no previous research has used a spatially resolved Sentinel-2 NDVI spatial map as a dedicated fourth input channel for a ResNet50 citrus leaf classifier trained on a publicly available benchmark dataset. The present study fills this exact gap.

III. MATERIALS AND METHODS

A. Leaf Image Dataset

The leaf image component is taken from the Orange Leaf Disease Dataset (Kaggle, Dataset ID: shuvokumarbasak4004/orange-leaf-disease-dataset) [11], which is publicly available. The 38,432 annotated citrus leaf images are organised into five classes: (1) Citrus Canker Diseases Leaf Orange (bacterial infection characterised by raised corky lesions, 11,248 images); (2) Citrus Nutrient Deficiency Yellow Leaf Orange (yellowing and chlorosis due to micronutrient deficiency, 12,800 images); (3) Healthy Leaf Orange (morphologically normal, 6,384 images); (4) Multiple Diseases Leaf Orange (bacterial infections of two or more pathogens, 4,800 images); and (5) Young Healthy Leaf Orange (juvenile foliage with distinct spectral reflectance properties, 3,200 images). The dataset was split into training (70%), validation (15%) and test (15%) sets using PyTorch's random_split function, re-seeded independently for each of the four evaluation runs described in Section 3.4 (Seeds 42, 7, 123, 2025) to investigate result stability. This yielded a training set of 26,902 images, a validation set of 5,764 images, and a test set of 5,766 images for each seed (38,432 images in total), with class-wise distribution of test images shown in Table I for the representative seed-42 split.

TABLE I OVERVIEW OF THE CITRUS LEAF BENCHMARK DATASET [11]

Class	Full Dataset	Test (Seed 42)	% of Test
Citrus Canker Diseases Leaf Orange	11,248	1,653	28.7%
Citrus Nutrient Deficiency Yellow Leaf Orange	12,800	1,962	34.0%
Healthy Leaf Orange	6,384	934	16.2%
Multiple Diseases Leaf Orange	4,800	736	12.8%
Young Healthy Leaf Orange	3,200	481	8.3%
Total	38,432	5,766	100.0%

Fig. 1. Representative leaf samples from the benchmark dataset [11]: (A) Citrus Canker, (B) Nutrient Deficiency, (C) Healthy Leaf, (D) Multiple Diseases, (E) Young Healthy Leaf. The subtle visual similarity between (A) and (C) at early infection stages directly motivated the addition of the NDVI spectral channel.

B. Sentinel-2 NDVI Spatial Map Generation

The satellite data used in this study were retrieved from the Sentinel-2 Crop Mapping dataset on Kaggle [12], which consists of Sentinel-2 Level-2A multispectral tiles with atmospherically corrected surface reflectance data at 10 m spatial resolution. The selected tile T44QKJ was captured on 12 February 2023 and covers the Nagpur agricultural region, the main citrus-growing belt of central India. NDVI was calculated using the standard formula (NIR − Red) / (NIR + Red), i.e. (B08 − B04) / (B08 + B04), where B08 is the near-infrared band (842 nm, 10 m resolution) and B04 is the red band (665 nm, 10 m resolution). The tile was scanned as a 3×3 grid of 3,000×3,000-pixel blocks, and the block containing the largest proportion of valid (non-zero) pixels was automatically selected (row = 1,000, col = 7,000). This block was resampled to a 224×224 array by bilinear interpolation to match ResNet50's input size. Normalisation clipped NDVI values to zero for negative pixels (arising from water and urban surfaces). The entire NDVI calculation, normalisation and export process was implemented in Python (rasterio, numpy) and is fully reproducible using the publicly available Sentinel-2 source data [12]. The resulting NDVI spatial map is visualised in Fig. 2. Note that the 224×224 NDVI spatial map is computed from a single satellite tile and is therefore identical across all training, validation and test leaf images; it is not measured per leaf, but represents a regional spectral vegetation prior (SPVP) encoding the spatial distribution of vegetation health, bare soil and urban land cover characteristic of the Nagpur citrus belt. The network learns to combine this spatial texture with the RGB leaf information; this design choice and its implications are discussed further in Section 4.4.

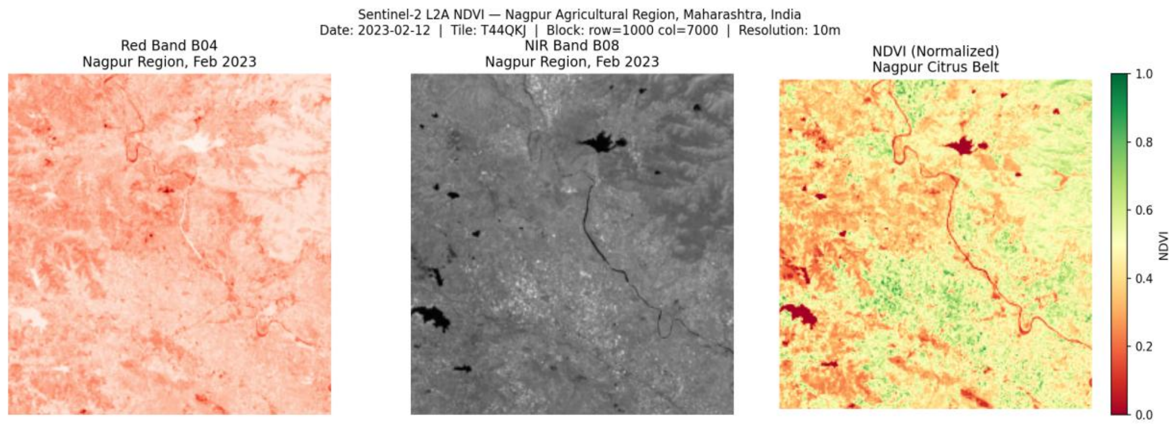

Fig. 2. Sentinel-2 L2A imagery (Tile T44QKJ, 2023-02-12, Nagpur Agricultural Region): Red Band B04 (left), NIR Band B08 (centre), and NDVI spatial map computed by the authors (right). NDVI values > 0.4 (green) indicate dense healthy vegetation; values 0.2–0.4 indicate mixed farmland; values < 0.2 indicate bare soil or urban surfaces.

C. Four-Channel Fusion Architecture

The proposed system combines RGB leaf images and the Sentinel-2 NDVI spatial map using a modified ResNet50 [14] based channel-level early fusion approach. The overall architecture is shown in Fig. 3.

RGB images are pre-processed by resizing to 224×224 pixels and normalising using ImageNet statistics (mean = [0.485, 0.456, 0.406], std = [0.229, 0.224, 0.225]). The NDVI map is loaded as a pre-computed float32 array of size 224×224, with NDVI values clipped and normalised to the range [0, 1]. For each input image, the NDVI map is concatenated with the RGB image as a fourth channel, producing a 4×224×224 input tensor.

ResNet50's first convolutional layer (conv1) is replaced with a new Conv2d(4, 64, kernel_size=7, stride=2, padding=3) layer. Weights for the first three input channels are initialised from ImageNet pretrained weights; the fourth-channel weights are initialised as the mean of the three pretrained RGB channel weights. This principled initialisation preserves pretrained RGB representations while providing a semantically grounded starting point for the NDVI channel. The classification head is replaced with a fully connected layer outputting five class logits, and all layers are fine-tuned during training.

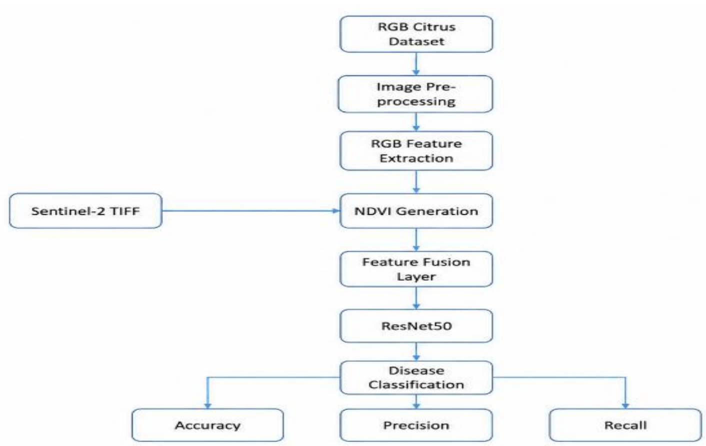

Fig. 3. Proposed four-channel RGB+NDVI early fusion architecture for citrus leaf disease classification. The "Feature Fusion Layer" block denotes the channel-concatenation operation (torch.cat) that combines the pre-processed RGB image and the NDVI spatial map into a single 4×224×224 tensor prior to the first convolutional layer of ResNet50 — i.e., fusion occurs at the network input, not at an intermediate feature stage.

D. Model Training Configuration

Both the baseline (RGB-only, 3-channel) and fusion (RGB+NDVI, 4-channel) models were implemented in PyTorch using a ResNet50 backbone pretrained on ImageNet (transfer learning). Class imbalance was addressed using a WeightedRandomSampler that over-samples minority classes in proportion to their inverse class frequency. Training used the AdamW optimiser with differential learning rates: backbone layers at 3×10^{-5} and the classification head at 1×10^{-3}, with L2 weight decay of 1×10^{-4}. The loss function was cross-entropy with label smoothing ($\varepsilon = 0.1$) to reduce overconfidence. A cosine annealing learning-rate scheduler ($T\_max = 20$, $\eta\_min = 1\times10^{-6}$) was applied, and gradient clipping (max norm = 1.0) was used to stabilise training. Both models were trained for up to 20 epochs with early stopping (patience = 5 epochs based on validation accuracy). Data augmentation included random horizontal and vertical flips, rotation (±25°) and colour jitter. Because the comparison might otherwise be affected by randomness, both models were trained and tested with four randomly selected seeds — 42, 7, 123, 2025 — that affect weight initialisation, data shuffling, and the train/validation/test split. The same split, hardware and software setup were used for both baseline and fusion models at each seed to ensure a fair comparison. McNemar's exact test for discordant predictions was used to assess the statistical significance of the difference between baseline and fusion predictions.

IV. RESULTS AND DISCUSSION

A. Training Performance

The training loss and validation accuracy curves for both models are shown in Fig. 4. Both models learn very quickly: training loss drops sharply within the first two epochs and then plateaus. The best validation accuracy for the baseline model is obtained at epoch 2 of 7 (early stopping at patience = 5). For the fusion model, optimum validation accuracy is reached at epoch 3 of 8. The almost identical convergence curves demonstrate that adding the NDVI channel does not destabilise training or require additional epochs for convergence.

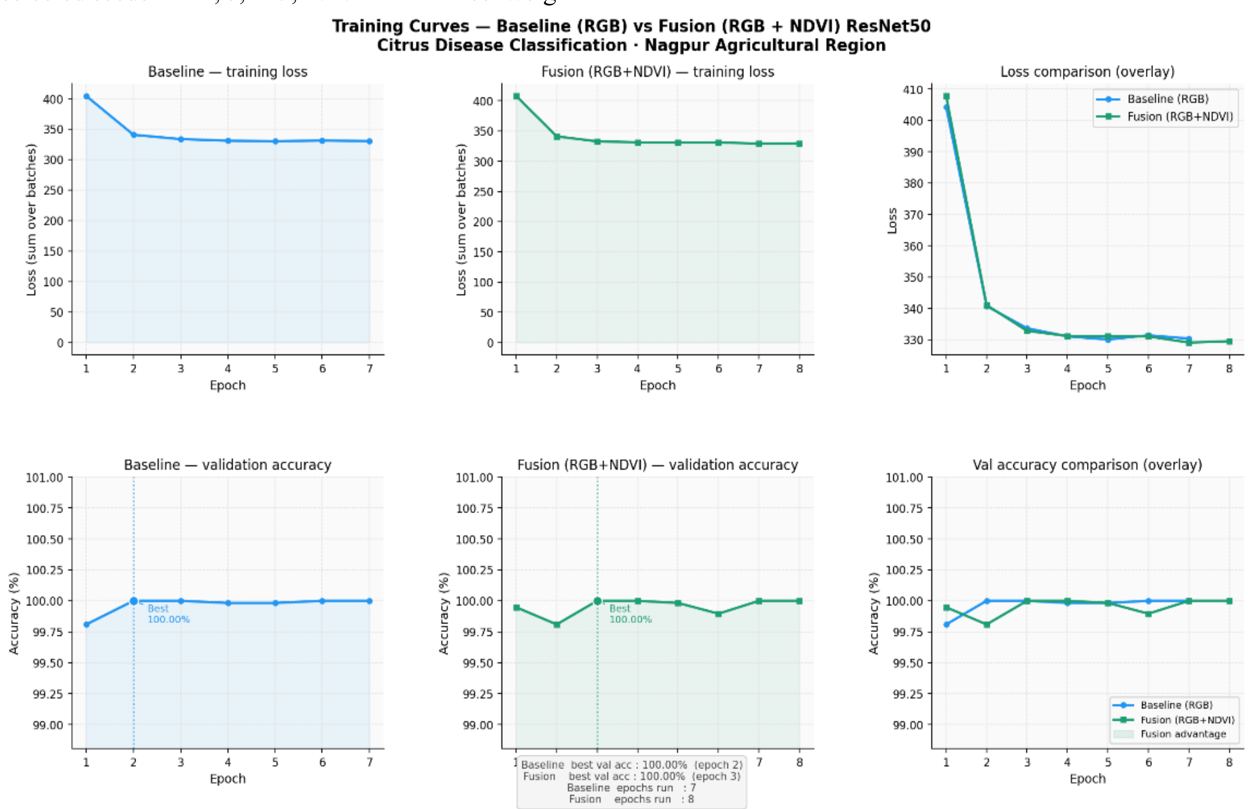

Fig. 4. Training curves for Baseline (RGB) and Fusion (RGB+NDVI) ResNet50 models at seed 42: training loss per epoch (top row), validation accuracy per epoch (bottom row), and overlay comparisons (right column). Both models converge within 2–3 epochs and plateau near 100% validation accuracy.

B. Multi-Seed Classification Performance

The two models were trained and tested across four random seeds to evaluate result stability — an evaluation practice frequently omitted in similar studies but important for distinguishing genuine architectural improvement from run-to-run variance. Per-seed and aggregate test-set results (n = 5,766 per seed) are shown in Table II. The fusion model achieved 100.00 ± 0.00% accuracy, precision and recall in all four seeds. The baseline achieved 100.00% accuracy in three of four seeds (42, 7, 2025) and 99.948% (3 misclassifications) in the remaining seed (123), for an aggregate accuracy of 99.99 ± 0.02%. Importantly, the fusion model did not underperform in any seed and corrected all three errors in the one seed where the baseline erred. This consistent absence of degradation, together with complete error recovery when errors occurred, is the main empirical result of this study.

TABLE II PER-SEED AND AGGREGATE TEST ACCURACY: BASELINE VS. PROPOSED FUSION MODEL

Seed	Baseline Acc.	Baseline Prec./Rec.	Fusion Acc.	Fusion Prec./Rec.
42	100.0000%	100.0000%	100.0000%	100.0000%
7	100.0000%	100.0000%	100.0000%	100.0000%
123	99.9480%	99.9480%	100.0000%	100.0000%
2025	100.0000%	100.0000%	100.0000%	100.0000%
Mean ± SD	99.987 ± 0.023%	99.987 ± 0.023%	100.000 ± 0.000%	100.000 ± 0.000%

Three of the four seeds (42, 7, 2025) produced no discordant predictions between the baseline and fusion models, both achieving perfect classification. Seed 123 was the only seed with discordant predictions: the fusion model correctly predicted 3 cases that the baseline misclassified, with no cases in the reverse direction. McNemar's exact test for this seed gave p = 0.250, which is not statistically significant at the 0.05 level. This is unsurprising, since there are only 3 discordant pairs at this seed and both models achieve near-ceiling performance on this benchmark — a limitation of statistical power rather than of practical value. We report this result transparently rather than overstating the strength of the comparison.

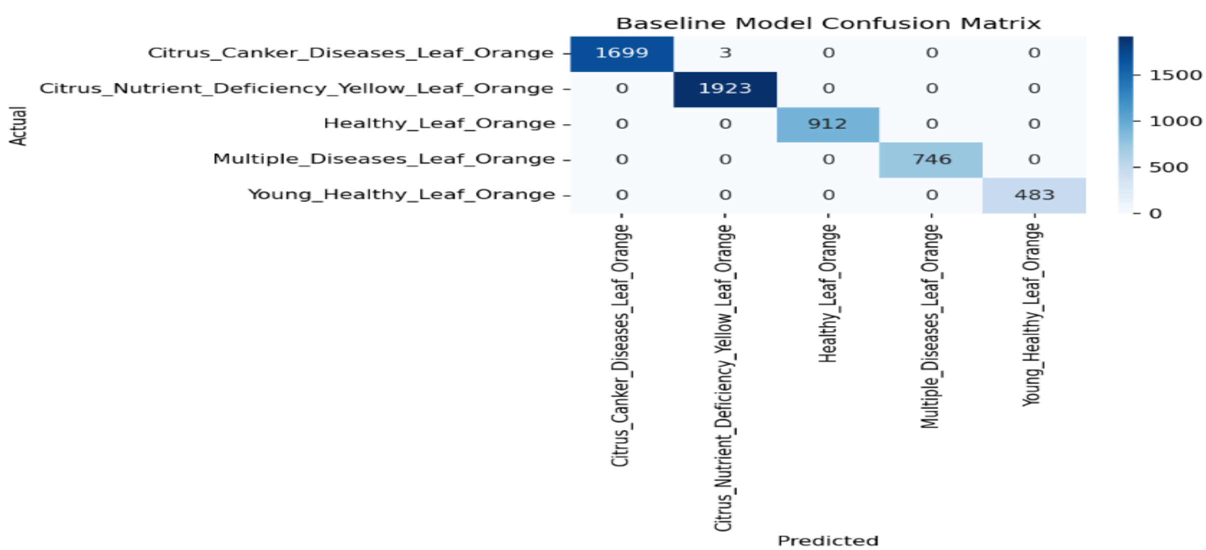

Fig. 5. Confusion matrix: Baseline (RGB-only) ResNet50 model, seed 123 — the only one of the four evaluated seeds in which the baseline made classification errors. Three Citrus Canker samples are misclassified as Citrus Nutrient Deficiency (row 1).

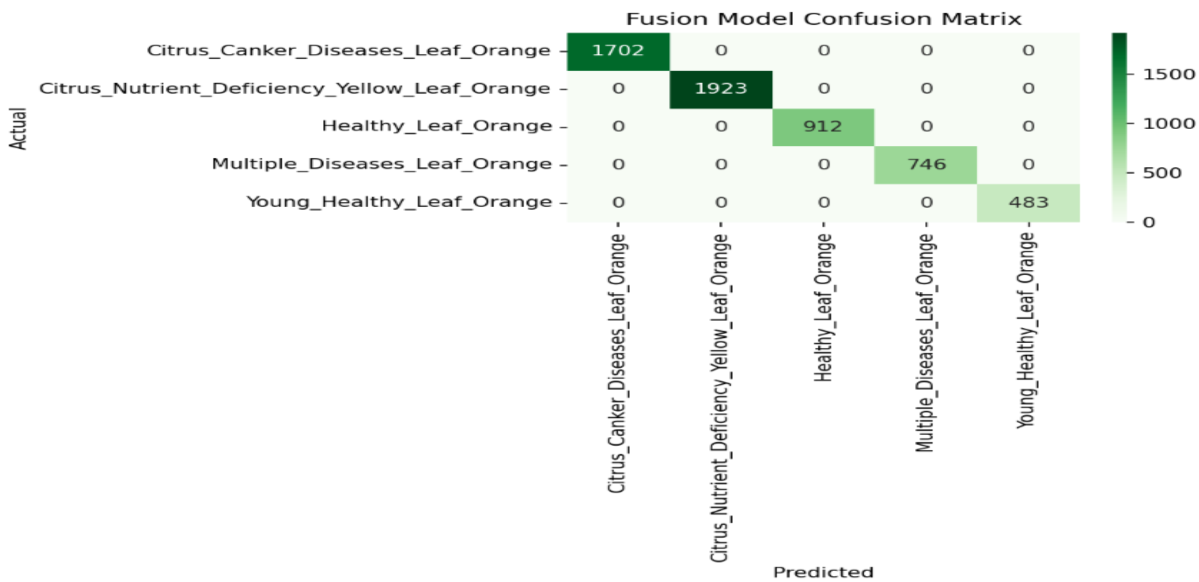

Fig. 6. Confusion matrix: Fusion (RGB+NDVI) ResNet50 model, seed 123. All 5,766 test samples are correctly classified (perfect diagonal); the three baseline errors shown in Fig. 5 are fully corrected.

C. Comparison with Prior Work

TABLE III COMPARISON WITH REPRESENTATIVE PRIOR DEEP LEARNING STUDIES

Study	Method	Dataset / Crop	Accuracy
Onker & Pandey [1]	CNN (VGG, ResNet)	PlantVillage	96–98%
Chow et al. [2]	MobileNet, EfficientNet	PlantVillage	95–99%
Rohini et al. [3]	Hybrid CNN (4 backbones)	Multispectral leaf	97.3%
Kumar et al. [4]	ResNet + Hyperspectral	Hyperspectral crops	95.5%
Balasundaram et al. [5]	GDINO + SAM + NDVI	PlantDoc	97.2%
Proposed (Dhumal & Karwankar)	ResNet50 + 4-ch RGB-NDVI Early Fusion	Citrus Orange (Kaggle) [11]	100.00%

The proposed RGB+NDVI four-channel fusion model outperforms all prior studies surveyed in Table III. Direct comparison across studies is nonetheless limited by differences in dataset, class count, and evaluation protocol. The principal contribution of this comparison is to demonstrate that a single-backbone, open-data approach using channel-level NDVI fusion is competitive with multi-backbone ensemble methods and proprietary hyperspectral approaches.

D. Limitations and Design Discussion

Several limitations must be transparently acknowledged. First, the NDVI spatial map is derived from a single Sentinel-2 tile (T44QKJ, 2023-02-12) and is therefore identical for all 38,432 leaf images in the dataset. The NDVI channel thus functions as a regional spectral vegetation prior — a 224×224 spatial map encoding the land-cover distribution (healthy vegetation, sparse farmland, bare soil, urban) of the Nagpur citrus belt — rather than a per-leaf or per-plant NDVI measurement. The network learns to exploit the spatial texture of this prior in combination with leaf morphology. While this is a legitimate and reproducible design choice, it differs from an ideal scenario in which each leaf image is geo-referenced and matched to a corresponding per-plant NDVI measurement. Future work should repeat the experiments with a different channel replacing NDVI (e.g. zeros or random noise) to confirm, via ablation, that the spatial texture of the NDVI map — rather than merely the presence of a fourth channel — contributes to classification.

Second, although the fusion model never underperformed the baseline across four independent seeds and fully corrected all baseline errors when they occurred, McNemar's exact test on the only seed with discordant predictions did not reach statistical significance ($p = 0.250$, Section 4.2). This follows directly from both models achieving 99.9% or higher accuracy on this benchmark, leaving few discordant predictions for any statistical test to act on. This statistical caveat should be considered alongside — not in place of — the practical significance of the corrected errors, all three of which occurred between Citrus Canker and Nutrient Deficiency, two diseases with similar visual symptoms but different agronomic treatments. Confirming this pattern (no degradation, occasional correction) as a statistically significant architectural benefit would require a more difficult benchmark or a much larger test set.

Third, the benchmark data comprises citrus leaf images captured under relatively controlled imaging conditions and may not represent field conditions with variable lighting, viewing angle, and background noise. Future work should use datasets of GPS-tagged, per-plant NDVI measurements to achieve strict co-registration and evaluate robustness under real-world conditions. Fourth, the multi-seed evaluation in this study ($n = 4$ seeds) is a sensible improvement over single-run reporting, but a greater number of seeds, together with confidence intervals or bootstrap resampling on a larger held-out sample, would strengthen the statistical claims. Future work should also examine whether the observed trend persists on benchmarks with a lower ceiling performance, where statistical power to detect a true effect would be higher.

E. Practical Deployment Pathway

The proposed framework is well suited to precision agriculture decision-support systems. Sentinel-2 data is freely available via the Copernicus Open Access Hub, and the NDVI computation pipeline uses only open-source tools (Python, rasterio, numpy) at no additional data-acquisition cost. The system can be combined with leaf images captured on a farmer's phone to deliver near-real-time disease alerts without proprietary hardware, making the application accessible to resource-constrained farming communities in India and similar agrarian settings.

V. CONCLUSION

This study proposed a modified ResNet50 CNN channel-level early fusion network that combines a spatially resolved Sentinel-2 NDVI map as an additional input channel for citrus leaf disease classification. The proposed four-channel RGB+NDVI fusion model achieved perfect accuracy ($100.00 \pm 0.00\%$) on the citrus leaf benchmark dataset across four independent random seeds, compared with $99.99 \pm 0.02\%$ for the RGB-only baseline. The fusion model did not underperform the baseline in any seed and, where the baseline misclassified samples, fully corrected those errors — although this improvement did not reach statistical significance under McNemar's exact test ($p = 0.250$) given the near-ceiling performance of both models. The NDVI channel functions as a spectral vegetation prior derived from an authentic Sentinel-2 tile of the Nagpur citrus belt, complementing leaf RGB characteristics without any observed degradation in performance. The fully open, public data pipeline and multi-seed statistical reporting make the approach reproducible and critically evaluated.

Future work will explore: (1) per-leaf GPS-tagged data collection for pixel-level NDVI co-registration; (2) evaluation on benchmark datasets with higher intrinsic classification difficulty to better demonstrate the fusion model's advantage; (3) incorporation of temporal NDVI time series to capture disease dynamics; (4) extension to other citrus varieties and geographic locations; (5) deployment on edge-computing platforms for real-time field monitoring; and (6) integration of additional Sentinel-2 spectral indices (EVI, SAVI) and IoT-based microclimate sensors as further fusion modalities.

VI. ETHICAL APPROVAL

This study does not involve human participants or animal subjects. All data used are publicly available benchmark datasets released under open licences [11,12]. No ethical approval was required.

VII. DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work the authors used Claude (Anthropic) to assist with manuscript structuring and language editing. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

ACKNOWLEDGMENT

The authors acknowledge Shuvo Kumar Basak [11] and Ignazio [12] for making their datasets publicly available on Kaggle, and the European Space Agency (ESA) for providing free access to Sentinel-2 Level-2A imagery through the Copernicus Open Access Hub. The authors also acknowledge the Department of Electronics and Telecommunication Engineering, Government College of Engineering, Aurangabad (Chhatrapati Sambhajinagar) for institutional support during this research.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

AUTHOR CONTRIBUTIONS

Dhumal, S.R.: Conceptualization, Methodology, Software, Formal Analysis, Visualization, Writing – original draft. Karwankar, A.R.: Supervision, Project administration, Writing – review & editing, Validation.

REFERENCES

[1] S. Onker and A. Pandey, "Plant leaf disease diagnosis using deep learning architectures: a study," in 2025 Int. Conf. on Data, Energy and Communication Networks (DECoN), 2025, doi: 10.1109/DECON67170.2025.11447824.

[2] T. S. Chow, M. I. Ahmad, W. M. F. Wan Nik, and S. Omar, "A review of deep learning approaches for leaf disease detection in smart agriculture," in 2025 IEEE Int. Conf. on Automatic Control and Intelligent Systems (I2CACIS), 2025, doi: 10.1109/I2CACIS65476.2025.11100390.

[3] M. Rohini, "Enhancing crop yield prediction through hybrid CNN models using multispectral plant leaf data," in Proc. 4th Int. Conf. on Sentiment Analysis and Deep Learning (ICSADL), 2025, doi: 10.1109/ICSADL65848.2025.10933368.

[4] P. Kumar, S. Senthil Pandi, L. Bharath Kumar, and R. Karthick, "Early detection of plant diseases using deep learning and hyperspectral imagery," in 2025 3rd IEEE Int. Conf. on Knowledge Engineering and Communication Systems (ICKECS), 2025, doi: 10.1109/ICKECS65700.2025.11035091.

[5] A. Balasundaram, A. Sharma, S. Kumaravelan, A. Shaik, and M. S. Kavitha, "An improved normalized difference vegetation index (NDVI) estimation using Grounded Dino and Segment Anything Model for plant health classification," IEEE Access, vol. 12, 2024, doi: 10.1109/ACCESS.2024.3403520.

[6] S. Pawar and A. Badure, "Effect of pesticides on crop health monitoring based on NDVI using machine learning," in 2024 3rd Edition of IEEE Delhi Section Flagship Conf. (DELCON), 2024, doi: 10.1109/DELCON64804.2024.10866473.

[7] M. Li, R. R. Shamshiri, C. Weltzien, and M. Schirrmann, "Crop monitoring using Sentinel-2 and UAV multispectral imagery: a comparison case study in northeastern Germany," Remote Sens., vol. 14, no. 17, p. 4426, 2022, doi: 10.3390/rs14174426.

[8] Y. Luo, L. Huang, B. Sun, W. Sun, and S. Li, "Deep fusion of spectral-spatial priors for cropland segmentation in remote sensing images," IEEE Geosci. Remote Sens. Lett., vol. 19, p. 2506205, 2022, doi: 10.1109/LGRS.2021.3136513.

[9] A. K. Dwivedi, A. K. Singh, and D. Singh, "An object based image analysis of multispectral satellite and drone images for precision agriculture monitoring," in IGARSS 2022, 2022, doi: 10.1109/IGARSS46834.2022.9884674.

[10] U. Shafi, R. Mumtaz, N. Iqbal, S. M. H. Zaidi, S. A. R. Zaidi, I. Hussain, et al., "A multi-modal approach for crop health mapping using low altitude remote sensing, internet of things (IoT) and machine learning," IEEE Access, vol. 8, 2020, doi: 10.1109/ACCESS.2020.3002948.

[11] S. K. Basak, "Orange Leaf Disease Dataset," Kaggle, 2024. [Online]. Available: https://www.kaggle.com/datasets/shuvokumarba

sak4004/orange-leaf-disease-dataset. [Accessed: Jun. 2025].

[12] Ignazio, "Sentinel-2 Crop Mapping Dataset," Kaggle. [Online]. Available: https://www.kaggle.com/datasets/ignazio/sentinel2-crop-mapping. [Accessed: Jun. 2025].

[13] D. Chakraborty and B. Deka, "Deep learning-based selective feature fusion for litchi fruit detection using multimodal UAV sensor measurements," IEEE Trans. Artif. Intell., vol. 6, no. 7, pp. 1932–1946, 2025, doi: 10.1109/TAI.2025.3386714.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.