Proceedings of the 6th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/88/20241694

Using computer vision to determine nutrient deficiency in winter crops

Alan Tai

Monta Vista High School, CA, USA

alantai19@gmail.com

Abstract. Winter wheat and winter rye often suffer from nutrient deficiencies due to variations in fertilizer usage, soil conditions, and other environmental factors. The timely and precise detection of these deficiencies, traditionally conducted through manual field inspection and soil testing, lacks scalability and often delays necessary interventions. To overcome these limitations, I conducted original research exploring the use of a Vision Transformer neural network, specifically the Swin Transformer V2, to classify crops into one of seven nutrient categories from UAV-based imagery. Two distinct, annotated datasets consisting of winter wheat (WW2020) and winter rye (WR2021) were used. Each dataset was subjected to an independent model training to account for visual differences in nutrient deficiencies between the crops. This approach leverages the advantages of the Vision Transformer, which has the capability to discern intricate patterns over expansive spatial regions. Top-1 accuracy was used to evaluate model predictions. A train-validation split yielded 67.7% accuracy for WW2020 and 68.2% for WR2021, while without a split, accuracy improved to 78.0% for WW2020 and 82.3% for WR2021. The Swin Transformer V2 shows promise for detecting nutrient deficiencies in winter crops. However, further fine-tuning and data collection are required to improve performance.

Keywords: computer vision, vision transformer, nutrient deficiencies, winter crops, Swin Transformer V2.

1. Introduction

Traditional methods for assessing crop nutrition are time-consuming and labor-intensive, relying on expert experience and chemical measurement which can delay necessary crop treatment. Additionally, these procedures can be subjective and invasive. Advancements in computer vision technology offer a promising avenue for crop nutrition detection, ensuring scalable real-time monitoring with rapid feedback, which can significantly optimize agricultural practices. These methods utilize image sensors to capture images of crops and subsequently process this data to extract crucial information about crop nutrition. For instance, prior research has shown that leaf spectral characteristics vary across different plants based on nitrogen nutrient levels [1]. Expanding the use of computer vision in agriculture can enable automated and intelligent management, diagnosing nutritional deficiencies with greater efficiency, reducing resource waste, and enhancing agricultural productivity [2].

Barbedo [3] finds that past research has primarily centered around the use of the following machine learning techniques: artificial neural networks (ANN), support vector machines (SVM), discriminant analysis techniques (DA), k-nearest neighbors (kNN), fuzzy logic, statistical classifiers, genetic

DOI: 10.54254/2755-2721/88/20241694

algorithms (GA), and random forests (RF). Deep learning is becoming the standard for image classification problems, and convolutional neural networks (CNNs) have been the most common choice.

In this research, I utilize an approach that harnesses a new architecture of ANNs instead of the common CNN architecture. Specifically, I use the Swin Transformer V2, a Vision Transformer with the ability to recognize intricate patterns over larger spatial regions, making it especially suited for challenges related to crop nutrient deficiency as compared to CNN-based methods.

2. Method

2.1. Dataset

As part of a challenge at the IEEE/CVF International Conference of Computer Vision (ICCV) 2023, organizers provided the Deep Nutrient Deficiency - Dikopshof - Winter Wheat and Winter Rye (DND-Diko-WWWR) dataset. DND-Diko-WWWR is a UAV-based RGB dataset consisting of 3,600 UAV-based RGB images, with 1,800 images of winter wheat harvested in 2020 (WW2020), and 1,800 images of winter rye harvested in 2021 (WR2021) captured at various dates. The plants were grown under seven distinct nutrient conditions and were annotated with corresponding nutrient statuses, as shown in figure 1. The nutrient classes included unfertilized, _PKCa, N_KCa, NP_Ca, NPK_, NPKCa, NPKCa+m+s, where "_" denotes the omission of the corresponding nutrient (N: nitrogen, P: phosphorus, K: potassium, Ca: lime), and +m+s represents additional application of mineral fertilizer and farmyard manure [5].

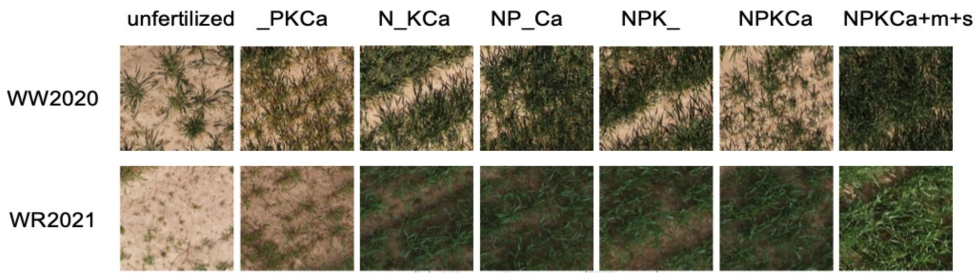

Figure 1. DND-Diko-WWWR Sample Images [4].

The organizers of the challenge originally divided the dataset into a trainval:test configuration of 74%:26% for both WW2020 and WR2021, where trainval refers to the subset of the dataset used during the model testing phase, and test refers to the subset of the dataset used for final evaluation of the model. Building upon this, I further segmented the trainval set to create dedicated training and validation subsets. Specifically, I opted for an approximately 80%:20% training:validation split within the trainval set. The rationale behind partitioning the trainval set was to enable the model's learning and evaluation phases. With 80% of the trainval data used for training, the model could adequately learn and adjust its parameters to accurately predict nutrient deficiencies from the crop images. The remaining 20% served as a validation set, assessing the model's performance during training to mitigate overfitting and ensure that the model generalizes well to new data. As depicted in figures 2 and 3, both WW2020 and WR2021 showcased an equitable representation of images for each class across training and validation subsets.

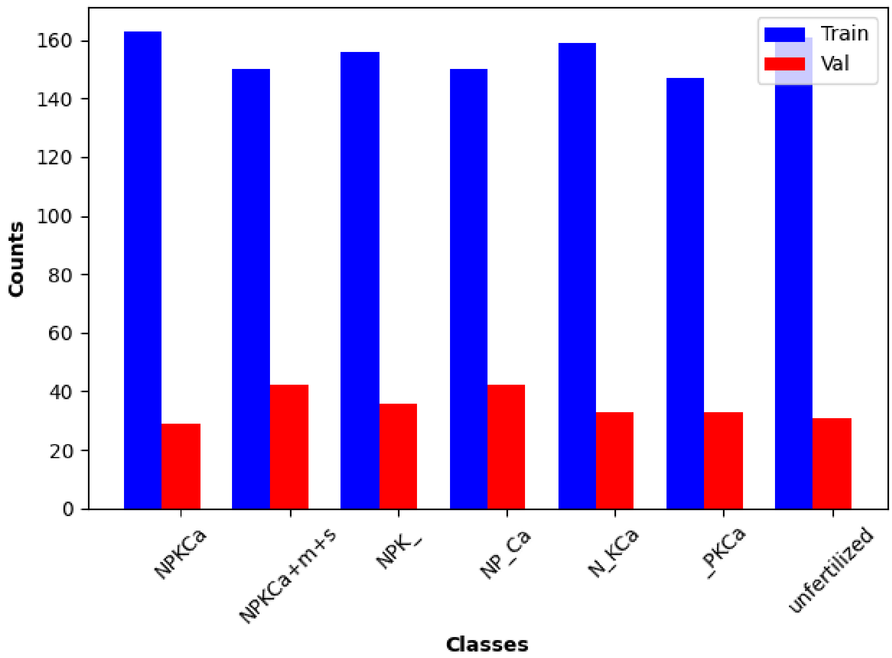

Figure 2. WW2020 Training-Validation Class Distribution.

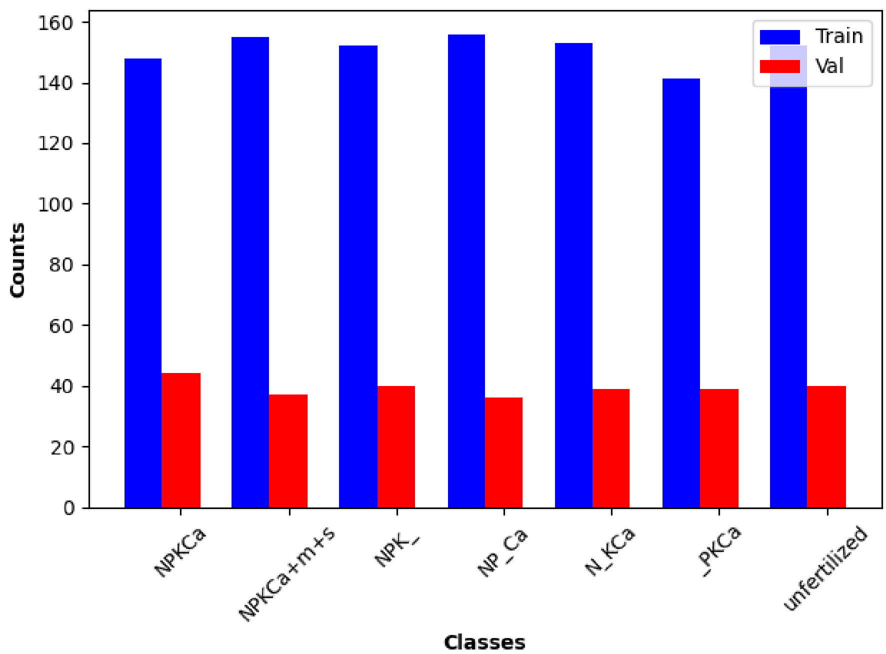

Figure 3. WR2021 Training-Validation Class Distribution.

The untouched 26% test set, as determined by the challenge organizers, was reserved for the ultimate evaluation of the model to simulate the model's real-world performance on unseen crop images, providing an indication of its progress for deployment in agricultural contexts.

Table 1. Distribution of Images Across Train, Validation, and Test Sets for WW2020 and WR2021.

	Number of Images			
	Training	Validation	Test	Total
WW2020	1086	246	468	1800
WR2021	1057	275	468	1800

2.2. Vision Transformer

Computer vision was utilized to address the challenges associated with traditional crop nutrient detection by leveraging the PyTorch framework to employ the Swin Transformer V2, a Vision Transformer (ViT) originating from the success of transformers in large-scale language models. ViTs segment input images into fixed-sized patches and utilize self-attention mechanisms, enabling them to discern long-range dependencies between visual elements [6]. This is applicable to crop imaging because the intricate patterns and variations in nutrient-deficient crops may manifest over larger spatial regions.

However, ViTs still face obstacles related to model capacity, training stability, and disparities in resolution. The Swin Transformer V2 is uniquely suited to overcome traditional limitations of ViTs, leveraging a residual-post-norm approach and a scaled cosine attention mechanism, emphasizing training stability. Specifically, the residual-post-norm approach means that normalization of the inputs is done after passing through a residual connection enabling the network to skip layers during training, and the scaled cosine attention mechanism uses the cosine function to measure similarities between input elements, scaling the similarity measure to ensure that it works efficiently across different input sizes and complexities [7].

2.3. Training

A structured training regimen was necessary to ensure the model effectively learned the patterns within the crop images. Each training session was over 50 epochs. The primary training routine began with an iterative process over predefined epochs. Each epoch, the model utilized batch processing to process the dataset in smaller batches, facilitating a more stable and effective learning process with the large DND-Diko-WWWR dataset.

Incorporating deep learning techniques, the model's forward pass generated scores that underwent a cross-entropy loss calculation against the true labels of the batch. This cross-entropy loss quantified the difference between predicted probability distributions and the actual distribution, as given by the following:

$$L(y, P) = -\sum_{c=1}^{C} y_c \log(P_c)$$

In the cross-entropy loss equation, C represents the number of classes in the classification task. y represents the true label vector and P is the predicted probability distribution. Specifically, y_c is the element of y corresponding to the c^{th} class. Similarly, P_c is the element of P corresponding to the c^{th} class, representing the predicted probability that the instance belongs to that class.

The loss was then backpropagated to adjust the model weights. The use of gradient accumulation allowed the optimizer, stochastic gradient descent (SGD), to update after processing a set of batches rather than every single batch, improving training stability in an environment with limited computational resources.

After the completion of every epoch, the validation set was used to assess the model's generalization capabilities. Key performance metrics were monitored, including batch loss and top-1 accuracy. Top-1 accuracy quantifies the frequency at which the model's primary prediction — its most confident classification — matches the actual label of the data. As training progressed, the decision of choosing the best model was made based on the highest accuracy achieved in the validation dataset across all epochs. This best performing model was saved for subsequent testing with the test dataset.

DOI: 10.54254/2755-2721/88/20241694

2.4. Accuracy

Measuring accuracy is crucial for this task to quantify the correctness of a model's predictions. In the context of this research, accuracy is indicative of the model's capability to correctly discern nutrient conditions in crop images. A high accuracy implies that a significant proportion of the total predictions made by the model align with the true conditions. Conversely, lower accuracy would suggest potential areas of improvement in the model or its training regimen.

The approach of top-1 accuracy was taken to calculate accuracy for WW2020 and WR2021. For each dataset, the predicted labels are compared with the true labels, and the number of correct predictions is determined. This count is then divided by the total number of predictions to compute the accuracy for each dataset.

3. Results

Submissions to the ICCV challenge were evaluated as top-1 accuracy, and overall rankings were done based on the average accuracy of WW2020 and WR2021 test set predictions.

The approach with a train-validation dataset split during training yielded 67.7% accuracy for WW2020 and 68.2% accuracy for WR2021. This resulted in a mean accuracy of approximately 67.9%.

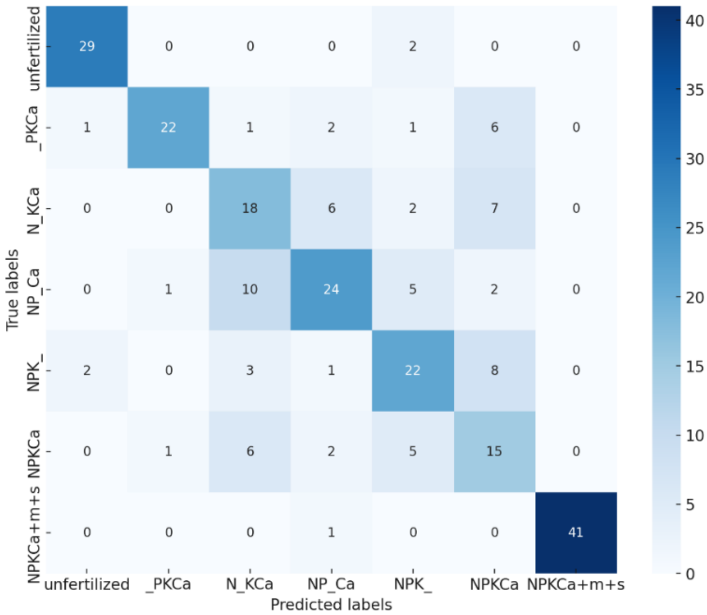

Figure 4. WW2020 Confusion Matrix Based on Train-validation Split Model Predictions on Validation Set.

DOI: 10.54254/2755-2721/88/20241694

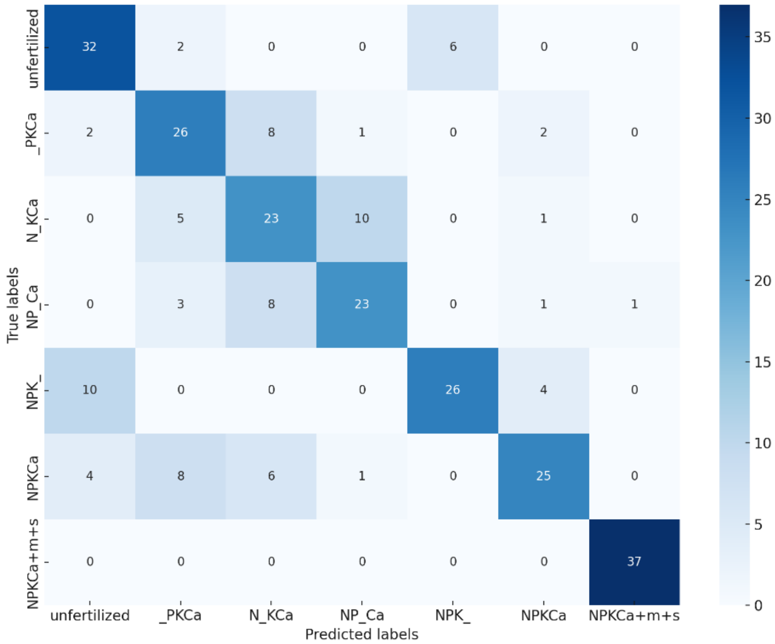

Figure 5. WR2021 Confusion Matrix Based on Train-validation Split Model Predictions on Validation Set.

An approach without a train-validation set split was also conducted, with training being done on the entire trainval set and the test set used for the evaluation of the model after each epoch. This approach resulted in 78.0% accuracy for WW2020 and 82.3% accuracy for WR2021, yielding a mean accuracy of approximately 80.1%.

4. Discussion

The findings from this research underscore the potential of computer vision, specifically using ViTs such as the Swin Transformer V2, to identify nutrient deficiencies in crops, such as winter wheat and winter rye. The substantial difference in accuracy between the two training approaches is surprising — 67.9% with a train-validation split and 80.1% without — suggesting that more data is needed to ensure a train-validation split results in improved performance, even though both the WW2020 and WR2021 training and validation sets had equitable representation of images for each class based on figures 2 and 3. By expanding the dataset to include a broader and more diverse range of crop conditions, a train-validation split would not compromise performance and ensure more robust generalization to real-world scenarios.

Additionally, based on figures 4 and 5, the ViT-based model correctly predicted the most nutrient-rich and completely unfertilized crops (class labels NPKCa+m+s and unfertilized, respectively) in terms of the number of images in the validation set predicted correctly. However, the model struggled with intermediate nutrient classes. For instance, the model failed to differentiate between N_KCa and NP_Ca in some instances on WW2020. As shown in figure 6, the two nutrient classes are likely difficult to distinguish for the model because of similar coloring and crop shape. Fine-tuning model weights could help to address this issue in future training.

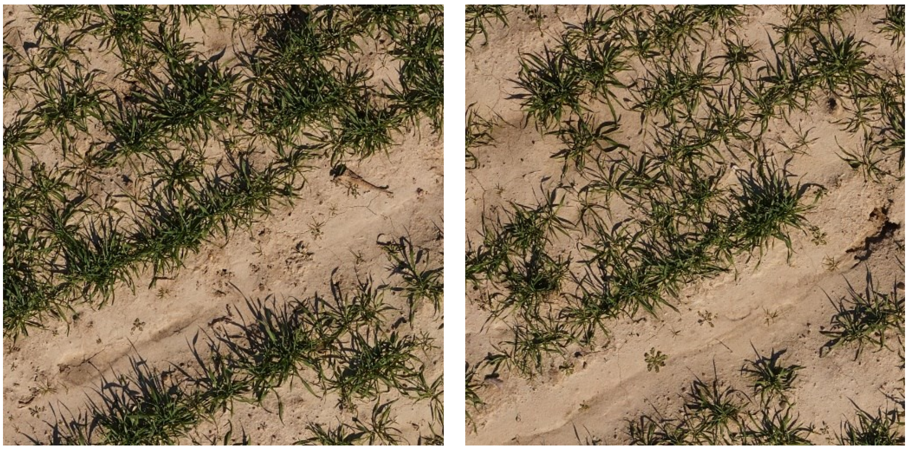

Figure 6. Images from WW2020 of N_KCa (left) and NP_Ca (right) [6].

Ultimately, the achieved accuracies represent a good starting point for applying computer vision to identify nutrient deficiencies in winter crops, with all approaches resulting in most predictions being accurate. This is a notable accomplishment considering that there are seven different nutrient classes depicted in the dataset — random guessing would have resulted in an accuracy of approximately 14.3%. The application of ViTs in assessing nutrient deficiencies in crops is thus promising, and further research should be centered on finding ways to fine-tune ViT-based models. The results also emphasize the importance of data quantity and diversity in achieving better model performance. Future research should explore methods to increase the data available to computer vision models, which may be done through techniques like data augmentation. These improvements would allow for expanded use of computer vision techniques in agriculture, ensuring targeted and efficient fertilizer application to bring about significant operational efficiencies and sustainability improvements in the industry.

Acknowledgments
I extend my deepest gratitude to Ross Greer, whose mentorship was indispensable throughout the course of this research. His profound insights, constant support, and feedback were pivotal in making this work possible. I would also like to thank the Gifted and Talented Institute of the Division of Undergraduate Education at the University of California, Irvine for providing the essential resources and feedback that greatly enhanced the quality of this research.

References

[1] Thomas, J. R., & Oerther, G. F. (1972). Estimating nitrogen content of sweet pepper leaves by reflectance measurements. *Agronomy Journal*, 64(1), 11-13.
[2] Peng, Q., & Xu, W. (2021). Crop nutrition and computer vision technology. *Wireless Personal Communications*, 117, 887-899.
[3] Barbedo, J. G. A. (2019). Detection of nutrition deficiencies in plants using proximal images and machine learning: A review. *Computers and Electronics in Agriculture*, 162, 482-492.
[4] Yi, J. (2023, June 16). CVPPA@ICCV'23: Image Classification of Nutrient Deficiencies in Win ter Wheat and Winter Rye. Retrieved August 15, 2023, from https://codalab.lisn.upsaclay.fr/c ompetitions/13833
[5] Yi, J. (2023, August 28). DND-Diko-WWWR. Retrieved August 15, 2023, from https://github.c om/jh-yi/DND-Diko-WWWR
[6] Jamil, S., Piran, M. J., & Kwon, O. J. (2023). A comprehensive survey of transformers for computer vision. *Drones*, 7(5), 287.
[7] Liu, Z., Hu, H., Lin, Y., Yao, Z., Xie, Z., Wei, Y., ... & Guo, B. (2022). Swin transformer v2: Scaling up capacity and resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 12009-12019).