

# Nouvelles approches de fusion de caractéristiques d'images multispectrales hétérogènes pour la détection de la jaunisse de la vigne

Shusong HUANG<sup>1</sup> Shurong ZHANG<sup>1</sup> Tanguy DAVAL<sup>1,2</sup> Eric PERRIN<sup>1</sup> Alban GOUPIL<sup>1</sup> Valeriu VRABIE<sup>1</sup>  
Marie-Laure PANON<sup>3</sup>

<sup>1</sup>CRESTIC ; Université de Reims Champagne-Ardenne, Moulin de la Housse, Chem. des Rouliers, 51100 Reims, France

<sup>2</sup>SEGULA Technologies - Reims, 5 Rue du Président Franklin Roosevelt, 51100 Reims, France

<sup>3</sup>Comité Champagne, 5 Rue Henri Martin, 51200 Épernay, France

shusong.huang@univ-reims.fr, marie-laure.panon@civc.fr

**Résumé** – Flavescence Dorée, l'une des jaunisses de la vigne, a un impact sévère sur la pérennité des vignobles. La détection de cette maladie à partir d'images multispectrales constitue un problème de classification, mais pose des défis. Dans ce travail, nous proposons de nouvelles approches pour améliorer la classification d'images multispectrales hétérogènes. Notre stratégie s'appuie sur une première fusion de caractéristiques des réseaux de neurones profonds et sur la mise en œuvre d'une seconde fusion de caractéristiques, la fonction softmax suivie d'un mécanisme d'attention par canal étant la meilleure approche. La fonction softmax utilise une normalisation exponentielle qui exploite la distribution de probabilité entre une carte de caractéristiques de groupe et une carte intergroupe afin d'optimiser leur fusion. Ensuite, le mécanisme d'attention par canal permet de capturer des caractéristiques complémentaires issues de différents canaux. Les résultats démontrent que la méthode proposée surpasse l'approche de l'état de l'art et d'autres possibilités de fusion et souligne le potentiel de cette stratégie pour une détection robuste de la jaunisse de la vigne.

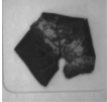
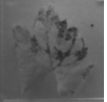
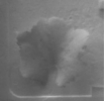
**Abstract** – Flavescence Dorée, one of the grapevine yellows diseases, has a severe impact on the sustainability of vineyards. Detecting this disease from multispectral images is a challenging classification problem. In this work, we propose new approaches to improve the classification of heterogeneous multispectral images. Our strategy is based on a first feature fusion of deep neural networks and the implementation of a second feature fusion, with the softmax function followed by a channel-based attention mechanism being the best approach. The softmax function uses an exponential normalization that exploits the probability distribution between a group feature map and an intergroup map to optimize their fusion. Then, the channel attention mechanism captures complementary features from different channels. The results demonstrate that the proposed method outperforms the state-of-the-art approach and other fusion possibilities, and underlines the potential of this strategy for robust detection of grapevine yellows.

## 1 Introduction

La Flavescence Dorée, l'une des jaunisses de la vigne, est une maladie grave de plus en plus répandue dans de nombreux pays qui peut entraîner une diminution significative de la production et de la qualité des vignobles [1, 2]. Il est essentiel de la détecter avant qu'elle s'étende dans les parcelles d'une année à une autre. La méthode traditionnelle d'identification, reposant sur l'expertise humaine et des analyses en laboratoire, présente des limites en termes d'efficacité et de précision. De plus, les symptômes de la maladie peuvent être confondus avec d'autres symptômes. L'imagerie multispectrale est une solution intéressante pour améliorer la détection des jaunisses de la vigne, comme l'ont montré les travaux de [3]. Une approche combinant la fusion de caractéristiques avec une analyse en composantes principales (ACP) appliquée à des images multispectrales a été proposée pour la détection de la maladie sur une même année [4]. D'autres outils s'appuyant sur des réseaux de neurones profonds ont également été testés [5, 6]. Cependant, ces derniers rencontrent les mêmes problèmes liés aux facteurs endogènes (vigueur de la vigne, etc.) et exogènes (notamment météo et phytotraitements) qui induisent d'importantes fluctuations au niveau des cartes de caractéristiques (features) extraites aboutissant à de fortes pertes de performances.

Des méthodes de fusion des caractéristiques, telles que la

TABLE 1 : Bandes Rouge et Red-edge d'images multispectrales acquises en 2021 et 2022 ; 'J' : Jaunisse, 'T' : Témoin

| Classe | 2021                                                                                 |                                                                                       | 2022                                                                                  |                                                                                       |
|--------|--------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|
|        | Rouge                                                                                | Red-edge                                                                              | Rouge                                                                                 | Red-edge                                                                              |
| J      |  |  |  |  |
| T      |  |  |  |  |

concaténation et l'addition, avec ou sans pondération, ont été introduites par [7, 8] puis des approches plus avancées basées sur l'attention en Deep Learning [9]. Cependant, la fusion des caractéristiques pour des images multispectrales hétérogènes reste un problème ouvert. En effet, comme l'illustre la Tableau 1 les images multispectrales mettent en évidence une variabilité importante des symptômes sur les images multispectrales, qui diffèrent selon les classes, les années et les bandes utilisées. Cette variabilité rend la combinaison des caractéristiques issues de ces bandes particulièrement complexe. Il apparaît donc essentiel d'approfondir l'étude des stratégies de fusion afin d'optimiser l'exploitation des informations contenues dans les

images multispectrales hétérogènes.

Dans ce contexte, nous proposons de développer une nouvelle méthode de fusion des caractéristiques profondes. Notre approche repose sur une seconde fusion via la fonction softmax, permettant une meilleure intégration des caractéristiques issues d'images multispectrales hétérogènes. Cette fusion peut être suivie d'un mécanisme d'attention au niveau des canaux, afin d'extraire des caractéristiques complémentaires provenant de différentes bandes spectrales. Cette stratégie innovante améliore significativement la fusion des informations par rapport aux approches existantes.

## 2 Méthodologie

Dans la Figure 1, nous explorons différentes approches pour fusionner des caractéristiques issues d'images acquises dans des bandes spectrales distinctes et proposons une nouvelle méthode. Celles-ci visent à enrichir l'information en combinant des caractéristiques d'images spectrales hétérogènes (architecture I), puis en exploitant des features complémentaires des canaux multiplexés (architecture II), et ensuite, y compris à l'aide d'un mécanisme d'attention sur les canaux (architecture III). Ce modèle repose sur une approche à deux niveaux : le niveau 1, qui réalise l'extraction de caractéristiques, et le niveau 2, qui réalise leur fusion.

Comme proposé dans [4], les images multispectrales sont dupliquées sur trois canaux et traitées indépendamment avec un même modèle appelé "pass model", EfficientNetB0 [10], par le premier niveau, voir l'architecture I dans la Figure 1. Les caractéristiques (A, B et C) sont extraites au niveau de la couche GN et fusionnées (par concaténation dans [4]) puis injectées au niveau de la couche GN+1 d'un autre modèle EfficientNetB0 appelé "cut model". Les modèles EfficientNetB0 sont pré-entraînés et sont affinés sur nos images dans le but de réaliser une classification sur  $k$  classes pour chacun des niveaux.

Par rapport à cette approche de l'état de l'art, nous proposons d'autres possibilités de fusionner les caractéristiques. De plus, nous proposons de fusionner les images multispectrales en créant l'équivalent d'une image RGB, mais en utilisant les mêmes 3 bandes spectrales à partir desquelles les caractéristiques A, B et C ont été extraites. Cela permet d'extraire de nouvelles caractéristiques de niveau 1, les caractéristiques D. Celles-ci peuvent être fusionnées dans un second temps avec les premières caractéristiques (fusion 2, cf. architectures II et III) en utilisant les mêmes possibilités de fusion.

Nous explorons de nouvelles stratégies pour fusionner les caractéristiques de niveaux 1 et 2. Nous avons d'abord exploré 3 méthodes de fusion des caractéristiques A, B et C afin d'obtenir la carte de caractéristiques notée  $\mathbf{f}_{1,c}$  en plus de la concaténation : le produit terme à terme, l'addition terme à terme, et la fusion Softmax. La fusion Softmax utilise une fonction exponentielle normalisée qui exploite la distribution de probabilité  $\mathbf{P}$  entre des caractéristiques, par exemple A, B et C pour le niveau 1. Les caractéristiques fusionnées sont transmises au 'cut model', qui sera réentraîné pour la classification en  $k$  classes.

Nous proposons d'intégrer davantage d'informations nouvelles issues d'images hétérogènes dans l'architecture II. Les caractéristiques  $\mathbf{F}_c$  combinent les caractéristiques  $\mathbf{f}_{1,c}$  et  $\mathbf{f}_{2,c}$

à l'aide de la fonction Softmax [11] s'écrivent de la façon suivante :

$$\mathbf{F}_c = \sum_{l=1}^n (\mathbf{f}_{l,c} \otimes \mathbf{P}_{l,c}), \quad (1)$$

où  $c$  désigne le canal,  $l$  la caractéristique allant de 1 à  $n = 2$ ,  $\mathbf{f}_{l,c} \in R^{C \times H \times W}$  la matrice de caractéristiques,  $C, H, W$  représentent respectivement les canaux, la hauteur et la largeur des caractéristiques,  $\otimes$  le produit élément par élément. Les probabilités de la fonction softmax sont calculées à l'aide de l'équation :

$$\mathbf{P}_{l,c} = (e^{\mathbf{f}_{l,c}} / \sum_{l=1}^n \sum_{c=1}^C \sum_{i=1}^H \sum_{j=1}^W e^{\mathbf{f}_{l,c}}), \quad (2)$$

avec  $\sum_{l=1}^n \mathbf{P}_{l,c} = 1$ ,  $i$  et  $j$  représentent les coordonnées spatiales du pixel de la caractéristique.

L'architecture II ajoute donc une seconde fusion entre les caractéristiques  $\mathbf{f}_{1,c}$  et  $\mathbf{f}_{2,c}$  en utilisant toujours les 4 mêmes méthodes de fusion décrites ci-dessus.

L'architecture III ajoute après la 2e fusion un mécanisme d'attention (CAM) qui est appliqué avec des poids  $\mathbf{w}(\mathbf{F})$  associés aux différentes caractéristiques fusionnées  $\mathbf{F}_c$ . Cela permet de prendre en compte des caractéristiques complémentaires des différents canaux afin de fournir la carte de caractéristiques finale, qui aura pour expression [9] :

$$\mathbf{F}'_c = \mathbf{F}_c \otimes \mathbf{w}(\mathbf{F}_c) = \mathbf{F}_c \otimes \sigma(\mathbf{G}(\mathbf{F}_c) \oplus \mathbf{L}(\mathbf{F}_c)), \quad (3)$$

où  $\mathbf{w}(\mathbf{F}_c) \in R^{C \times H \times W}$  représente les poids de l'attention et  $\sigma$  est une unité linéaire rectifiée (ReLU). Le contexte global du canal  $\mathbf{G}(\mathbf{F}_c)$  comporte une opération de pooling global moyen, une couche de convolution pixel par pixel, une fonction ReLU et des couches de convolution pixel par pixel. Le contexte local du canal  $\mathbf{L}(\mathbf{F}_c)$  comporte une couche de convolution pixel par pixel, une fonction ReLU et des couches de convolution pixel par pixel, comme il est précisé dans [9].  $\oplus$  représente le broadcasting addition.

## 3 Application

### 3.1 Images multispectrales

Des feuilles de vigne ont été prélevées durant quatre années dans le vignoble expérimental du Comité Champagne à Épernay (France) pendant la période des vendanges. Seules les feuilles correspondant à deux classes, Jaunisse et plantes saines (Témoin) ont été utilisées dans cette étude. Les données contiennent 44.2%, 55.8% de feuilles Jaunisse et de feuilles témoins respectivement. Le protocole de mesure consiste à placer quatre feuilles sur une plaque blanche en polystyrène afin de disposer d'un arrière-plan simple et facilitant la segmentation (une feuille par image). Les acquisitions des images multispectrales ont été réalisées à l'aide d'une caméra multispectrale DJI. Cette caméra fournit des images multispectrales dans cinq bandes Bleu, Vert, Rouge, Red-edge et NIR, ainsi que des images RGB avec une résolution spatiale de  $1\,000 \times 1\,000$  pixels. Les images RGB ont été utilisées pour labelliser les images multispectrales par des experts du Comité Champagne. Le nombre d'images par classe est fourni dans le tableau 2.

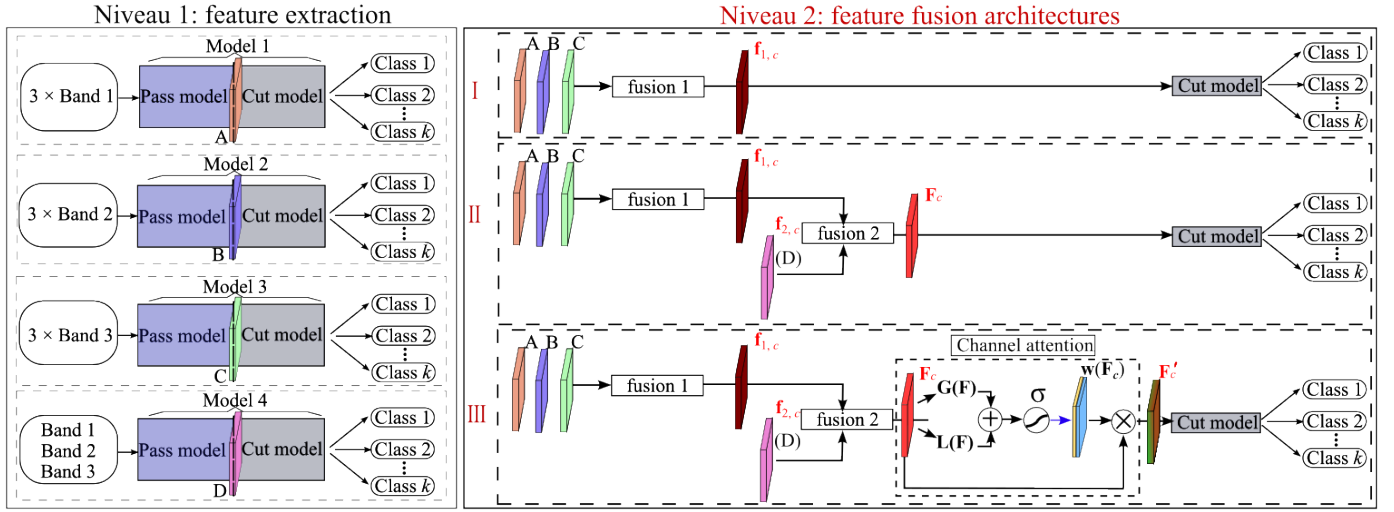


FIGURE 1 : Architectures globales proposées pour la fusion profonde

TABLE 2 : Nombre d'images multispectrales (4 bandes)

| Classe\Année | 2021-2022 | 2023  | 2024  |
|--------------|-----------|-------|-------|
| Jaunisse     | 1,380     | 1,532 | 2,420 |
| Témoin       | 2,336     | 1,840 | 2,560 |

### 3.2 Préparation du jeu de données

Les images multispectrales capturées avec la caméra DJI présentent des décalages entre elles étant acquises avec des capteurs différents (phénomène de parallaxe). Un prétraitement a été réalisé afin de les aligner les unes par rapport aux autres. Néanmoins, étant donné que pour des acquisitions sur le terrain, en fonction de l'orientation de la caméra, des décalages variables entre les images peuvent apparaître, et afin d'étudier la robustesse des architectures par rapport à ces conditions complexes, les images multispectrales ont été légèrement décalées de manière aléatoire jusqu'à 5% par rapport aux images alignées suivant les directions verticale et horizontale, afin de garantir que les feuilles restent dans le cadre. Nous avons également redimensionné toutes les images multispectrales à une dimension de  $224 \times 224$  pour satisfaire la condition requise en entrée du modèle CNN que nous choisissons, ici l'EfficientNetB0.

### 3.3 Entraînement et tests des architectures

80% des données de 2021 et 2022 ont été utilisées pour l'entraînement. Trois tests différents avec des niveaux de difficulté croissants ont été mis en place.

**Test 1 :** Les 20% restants des données de 2021 et 2022 ont été utilisées pour valider le modèle.

**Test 2 :** Des données collectées en 2023 ont été utilisées pour évaluer la robustesse du modèle au fil des années.

**Test 3 :** Des données collectées en 2024 ont également été utilisées pour évaluer la robustesse des modèles.

### 3.4 Entraînement de modèles

Nous avons implémenté nos modèles à l'aide de la librairie TensorFlow 2.17.0 avec 2 GPU. Pour le niveau 1, les modèles 1 à 4 consistent en un réseau EfficientNetB0, dans lequel, après

le dernier average pooling, nous ajoutons 3 couches entièrement connectées (100, 10, 2) avec des fonctions d'activation ReLU pour les deux premières couches et une activation softmax pour la dernière couche. Les modèles 1 à 4 sont fine-tunés sur les 100 dernières couches et itérés sur 100 Epochs. Un optimiseur Adam avec un taux d'apprentissage qui décroît continuellement de  $10^{-3}$  à  $10^{-4}$  a été choisi. Un batch size de 32 a été sélectionné. Pour les modèles du niveau 1, l'extraction des caractéristiques est réalisée au niveau de la couche GN=23.

### 3.5 Résultats

Le Tableau 3 présente les résultats obtenus sans fusion de caractéristiques au niveau 1 en utilisant les trois canaux séparés puis en combinant les canaux Bleu-Vert-Rouge et Vert-Rouge-Red-edge. Il s'agit tout simplement de modèles EfficientNetB0 fine-tunés pour une tâche de classification à 2 classes sur les données d'entrée respectives. Les résultats montrent que la bande rouge fournit une meilleure précision que les autres bandes. De même, on peut constater que les images combinées Vert-Rouge-Red-edge permettent d'améliorer la précision par rapport aux images Bleu-Vert-Rouge, car la bande Red-edge contient des informations plus pertinentes que la bande bleue. Ces premiers résultats montrent qu'il est plus intéressant d'utiliser des images multispectrales plutôt que celles issues d'une caméra RGB classique, ces bandes ayant donc été utilisées par la suite.

TABLE 3 : Moyenne des précisions de niveau 1 pour chaque bande et pour la combinaison de trois bandes Bleu-Vert-Rouge (B-V-Ro) et Vert-Rouge-Red-edge (V-Ro-Re)

| Type       | Band         | Test 1 | Test 2 | Test 3 |
|------------|--------------|--------|--------|--------|
| Uniquement | 3 × Bleu     | 0.9304 | 0.7674 | 0.8324 |
|            | 3 × Vert     | 0.9839 | 0.8398 | 0.6285 |
|            | 3 × Rouge    | 0.9465 | 0.8505 | 0.8324 |
|            | 3 × Red-edge | 0.9625 | 0.8149 | 0.8157 |
| Combinée   | B-V-Ro       | 0.9893 | 0.7758 | 0.7179 |
|            | V-Ro-Re      | 0.9893 | 0.8303 | 0.8127 |

Pour évaluer les stratégies de fusion du niveau 2, nous avons calculé la précision (accuracy) et le score F1 moyen en utilisant les bandes Vert, Rouge, Red-edge et l'image combinée V-Ro-Re. Les résultats sont présentés dans le Tableau 4. Notre

TABLE 4 : Moyennes des précisions et F1 score des méthodes de fusion

| Arch. | Stratégie de fusion niveau 1 (+ niveau 2) | Test 1        |               | Test 2        |               | Test 3        |               |
|-------|-------------------------------------------|---------------|---------------|---------------|---------------|---------------|---------------|
|       |                                           | Accuracy      | F1 score      | Accuracy      | F1 score      | Accuracy      | F1 score      |
| I     | Concatenation [4]                         | 0.9786        | 0.9770        | 0.8185        | 0.8178        | 0.7574        | 0.7567        |
|       | Dot product                               | 0.9839        | 0.9885        | 0.8351        | 0.8297        | 0.7703        | 0.7662        |
|       | Somme                                     | 0.9947        | 0.9943        | 0.8636        | 0.8619        | 0.8168        | 0.8160        |
|       | Softmax                                   | 0.9893        | 0.9885        | 0.8540        | 0.8506        | <b>0.8891</b> | <b>0.8898</b> |
| II    | Somme + Concatenation                     | 0.9679        | 0.9656        | 0.7900        | 0.7899        | 0.8811        | 0.8811        |
|       | Somme + Dot product                       | 0.9947        | 0.9943        | 0.7782        | 0.7773        | 0.7020        | 0.6829        |
|       | Somme + Somme                             | 0.9947        | 0.9943        | 0.8719        | 0.8710        | 0.8281        | 0.8272        |
|       | Softmax + Softmax                         | 0.9947        | 0.9943        | 0.8505        | 0.8494        | 0.8402        | 0.8398        |
|       | Softmax + Somme                           | <b>1.0</b>    | <b>1.0</b>    | 0.8695        | 0.8658        | <b>0.8811</b> | <b>0.8811</b> |
|       | Somme + Softmax                           | <b>1.0</b>    | <b>1.0</b>    | <b>0.8743</b> | <b>0.8730</b> | 0.8072        | 0.8040        |
| III   | Somme + Concatenation + CAM               | 0.9733        | 0.9710        | 0.6762        | 0.6734        | 0.7574        | 0.7529        |
|       | Somme + Dot product + CAM                 | 0.9733        | 0.9712        | 0.7806        | 0.7803        | 0.8530        | 0.8525        |
|       | Somme + Somme + CAM                       | 0.9840        | 0.9828        | 0.8375        | 0.8350        | 0.8466        | 0.8465        |
|       | Softmax + Softmax + CAM                   | 0.9840        | 0.9827        | 0.8505        | 0.8450        | 0.8779        | 0.8759        |
|       | Softmax + Somme + CAM                     | 0.9947        | 0.9942        | 0.8482        | 0.8461        | 0.7671        | 0.7661        |
|       | Somme + Attention (embedded) CAM [9]      | 0.9786        | 0.9769        | 0.8565        | 0.8552        | 0.8337        | 0.8303        |
|       | Somme + Softmax + CAM                     | <b>0.9946</b> | <b>0.9957</b> | <b>0.8647</b> | <b>0.8810</b> | <b>0.8859</b> | <b>0.8841</b> |

méthode permet d’obtenir des prédictions parfaitement exactes (100%) pour la même année. Il est évident que nos modèles de fusion profonds améliorent considérablement la précision et le score F1 moyen par rapport à des méthodes simples sans fusion mais aussi par rapport à des modèles de l’état de l’art, [4] et [9].

## 4 Conclusion

Nous avons proposé un nouveau modèle de fusion profond de caractéristiques pour la détection de la jaunisse de la vigne à partir d’images multispectrales hétérogènes. Cette méthode s’est révélée particulièrement efficace pour la tâche de classification binaire. Les valeurs de précision (accuracy) et le score F1 montrent une amélioration significative par rapport à des méthodes de l’état de l’art. Notre travail ouvre de nouvelles perspectives pour une détection robuste. Parmi les pistes de recherche futures, nous envisageons des tâches d’augmentation des données, l’utilisation de bandes multispectrales supplémentaires ainsi que l’exploration de techniques d’embedding et de modèles basés sur des transformers.

## Références

- [1] M Maixner. Grapevine yellows—current developments and unsolved questions. In *Extended Abstracts 15th Meeting ICVG*, 2006.
- [2] G Belli, PA Bianco, and M Conti. Grapevine yellows in Italy : past, present and future. *Journal of Plant Pathology*, pages 303–326, 2010.
- [3] Johanna Albetis de la Cruz. Potentiel des images multispectrales acquises par drone dans la détection des zones infectées par la flavescence dorée de la vigne. *Thèse de doctorat, l’Université Paul Sabatier - Toulouse III*, 2018.
- [4] Guillaume Heller, Eric Perrin, Valeriu Vrabie, Cedric Dusart, Marie-Laure Panon, Marie Loyaux, and Solen Le Roux. Multisource neural network feature map fusion : An efficient strategy to detect plant diseases. *Intelligent Systems with Applications*, 19 :200264, 2023.
- [5] Justine Boulent, Pierre-Luc St-Charles, Samuel Foucher, and Jérôme Théau. Automatic detection of flavescence dorée symptoms across white grapevine varieties using deep learning. *Frontiers in Artificial Intelligence*, 3 :564878, 2020.
- [6] Antonios Morellos, Xanthoula Eirini Pantazi, Charalampos Paraskevas, and Dimitrios Moshou. Comparison of deep neural networks in detecting field grapevine diseases using transfer learning. *Remote sensing*, 14(18) :4648, 2022.
- [7] Yu Zhang, Cong Phuoc Huynh, and King Nghi Ngan. Feature fusion with predictive weighting for spectral image classification and segmentation. *IEEE Transactions on Geoscience and Remote Sensing*, 57(9) :6792–6807, 2019.
- [8] Xijian Fan, Peng Luo, Yuen Mu, Rui Zhou, Tardi Tjahjadi, and Yi Ren. Leaf image based plant disease identification using transfer learning and feature fusion. *Computers and Electronics in agriculture*, 196 :106892, 2022.
- [9] Yimian Dai, Fabian Gieseke, Stefan Oehmcke, Yiquan Wu, and Kobus Barnard. Attentional feature fusion. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 3560–3569, 2021.
- [10] Junde Chen, Jinxiu Chen, Defu Zhang, Yuandong Sun, and Yaser Ahangari Nanehkaran. Using deep transfer learning for image-based plant disease identification. *Computers and electronics in agriculture*, 173 :105393, 2020.
- [11] Xingqun Qi, Tianhui Wang, and Jiaming Liu. Comparison of support vector machine and softmax classifiers in computer vision. In *2017 Second International Conference on Mechanical, Control and Computer Engineering (ICMCCE)*, pages 151–155. IEEE, 2017.