A comprehensive review of 3D convolutional neural network-based classification techniques of diseased and defective crops using non-UAV-based hyperspectral images

Nooshin Noshiri [a,*], Michael A. Beck [a], Christopher P. Bidinosti [b], Christopher J. Henry [a,*]

[a] Applied Computer Science Department, The University of Winnipeg 515 Portage Ave, Winnipeg, R3B 2E9, Manitoba, Canada
[b] Physics Department, The University of Winnipeg 515 Portage Ave, Winnipeg, R3B 2E9, Manitoba, Canada

ARTICLE INFO

ABSTRACT

Hyperspectral imaging (HSI) is a non-destructive and contactless technology that provides valuable information about the structure and composition of an object. It has the ability to capture detailed information about the chemical and physical properties of agricultural crops. Due to its wide spectral range, compared with multispectral- or RGB-based imaging methods, HSI can be a more effective tool for monitoring crop health and productivity. With the advent of this imaging tool in agrotechnology, researchers can more accurately address issues related to the detection of diseased and defective crops in the agriculture industry. This allows to implement the most suitable and accurate farming solutions, such as irrigation and fertilization, before crops enter a damaged and difficult-to-recover phase of growth in the field. While HSI provides valuable insights into the object under investigation, the limited number of HSI datasets for crop evaluation presently poses a bottleneck. Dealing with the curse of dimensionality presents another challenge due to the abundance of spectral and spatial information in each hyperspectral cube. State-of-the-art methods based on 1D and 2D convolutional neural networks (CNNs) struggle to efficiently extract spectral and spatial information. On the other hand, 3D-CNN-based models have shown significant promise in achieving better classification and detection results by leveraging spectral and spatial features simultaneously. Despite the apparent benefits of 3D-CNN-based models, their usage for classification purposes in this area of research has remained limited. This paper seeks to address this gap by reviewing 3D-CNN-based architectures and the typical deep learning pipeline, including preprocessing and visualization of results, for the classification of hyperspectral images of diseased and defective crops. Furthermore, we discuss open research areas and challenges when utilizing 3D-CNNs with HSI data.

1. Introduction

Plant diseases pose significant threats to global food production, with potential yield losses of up to 30% and substantial economic impact [56]. This can have a devastating impact on farmers and communities, particularly in low-income countries where access to food is already challenging. Precision agriculture and hyperspectral imaging (HSI) offer promising solutions for preventing crop damage and losses, ultimately contributing to efforts to promote sustainability and reduce the impact of diseases on food production.

HSI, also referred to as imaging spectrometry, combines two distinct technologies, imaging and spectroscopy, to provide both spatial and spectral information, simultaneously. Spectral information can provide rich information about biochemical and biophysical attributes of the agricultural crops. This is due to the higher spectral resolution of hyperspectral sensors compared to multispectral and RGB ones. As a result, this feature can lead to better discrimination of objects of similar colors, higher accuracy in complex classifications, and the ability to predict chemical composition and provide information about the interior of an object [67].

However, the interpretation of spectral data can be complex, especially when analyzing and comparing multiple samples over extended periods. One approach to simplify spectral analysis is the usage of spectral indices. Spectral indices are mathematical expressions that combine
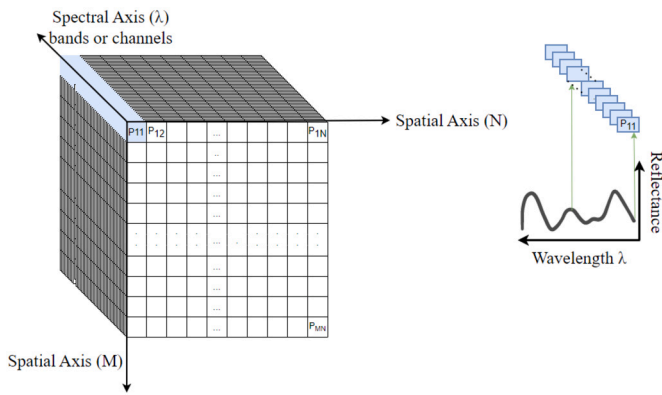
Fig. 1. The hyperspectral cube (adapted from Tarabalka et al. [74] with modification). It is a three-dimensional array where each pixel represents a spectrum containing a range of wavelengths. This spectrum can act as a fingerprint and provides information about biophysical and biochemical characteristics of the imaged object.

several spectral bands into a single value, providing an easier representation of the data. For example, the Normalized Difference Vegetation Index (NDVI) computes the ratio between Near Infrared (NIR) and Red (R) bands of hyperspectral channels as follows:

$$NDVI = \frac{NIR - R}{NIR + R} \tag{1}$$

With the help of spectral indices, one can effectively identify trends and changes in the data without requiring an in-depth comprehension of the underlying scientific principles governing spectral data, thus enabling a simpler data analysis, enhancement of features, standardization, comparability, and calibration of data.

In the agricultural industry, two common spectral indices are the already mentioned NDVI and the Green Chlorophyll Index (GCI). The former is used to monitor vegetation growth and health and the latter quantifies the amount of chlorophyll in plants. Further indices have been defined, usually in the context of remote sensing, to support research for example on agriculture, soil, vegetation, water and forestry. A comprehensive database of spectral indices that is searchable by application area and hyperspectral sensor is provided in Henrich et al. [24].

From a data-perspective, a hyperspectral image is a stack of images, known as a hyperspectral cube or a data cube. Each image of this cube represents the response of the imager to one of the distinct hyperspectral channels [5]. This is illustrated in Fig. 1. It shows a 3D data cube P with dimensions $M \times N \times \lambda$, where M and N represent the axes of spatial information and λ represents the spectral dimension [74]. In the hyperspectral cube, each pixel, given by its spatial coordinates, is a vector of length λ that indicates the reflected radiation of a specific part of the object.

The high-dimensionality of this data cube poses a challenge to traditional machine learning approaches, resulting in reduced accuracy due to their inability to extract complex features. Moreover, the performance of these approaches heavily depends on manual feature engineering. Convolutional Neural Networks (CNN) have been proven to achieve high classification accuracies in image classification tasks and to work well with the high-dimensionality of HSI data.

In this paper, we present a comprehensive review of 3D-CNN-based models utilized in the classification of non-UAV-based hyperspectral images of diseased and defective crops. This review is intended to assist computer vision experts and agriculture-domain researchers seeking to address HSI classification tasks for crops under stress.

The paper is organized as follows. In the following Section, we outline the investigation protocol used in this review. In Section 3, we briefly describe the structure of CNNs and their most important concepts. Following the typical data pipeline associated with CNNs, Sections 4 to 7 review data preprocessing, band and feature selection,

network architecture design, and data visualization. This provides a convenient overview of its individual steps for plant classification problems using 3D-CNNs. Finally, Section 8 highlights the research gaps and limitations associated with the application of 3D-CNNs for HSI data classification.

2. Search methodology

A systematic search was conducted by accessing scholarly publications through Google Scholar search engine. To optimize the search results, specific keywords were employed in the advanced search section, resulting in a refined list of articles. The selected search terms were different combinations of "hyperspectral", "disease", "detection", "identification", "diagnosis", "plant", "crop", "stress", "3D CNN", "3 dimensional CNN", "three dimensional CNN", utilizing the Boolean operators, AND and OR. Around 2,000 records were investigated, however, to ensure the relevance of the articles, the abstracts of the retrieved papers were thoroughly assessed to confirm coherence with the title of this research. Moreover, screening criteria were implemented, including the removal of non-English papers, to ensure an accessible and high-quality selection of articles. The results of this comprehensive study are provided in Table 1 and Table 2 (see Section 6) and contain papers from 2015 up to February of 2023.

This review examines various applications of 3D-CNN-based models in detecting and classifying diseases in agricultural crops, including charcoal rot in soybeans [45], mold in peanuts and strawberries [42,33], bacterial leaf blight (BLB) in rice [10], grapevine vein-clearing virus (GVCV) in grapevines [47], and potato late blight (PLB) in potatoes [51]. Moreover, the review explores the use of 3D-CNN-based architectures for identifying specific defects in crops, such as decay in blueberries [52], bruise and brown spots in fruits [50,28], heat stress in rice [19], as well as black, fermented, shell, and broken coffee defects in beans [13]. By harnessing the power of 3D-CNN-based models, we can effectively address challenges of classifying diseased and defective using hyperspectral images. This can result in the preservation of product quality, prevention of yield losses, and ensuring food safety standards.

3. CNN structure and concepts

A CNN is comprised of a series of layers each consisting of several neurons. As shown in Fig. 2, each layer is the input for the next layer in the network. Key building blocks of a CNN are convolution layers, detector layers, and pooling layers. A convolution layer uses convolutional kernels to extract low-dimensional features from the input data while preserving the spatial relationship between the input data pixels. Fig. 3 depicts the movement directions of 1D (spectral), 2D (spatial), and 3D (spatial-spectral) kernels within the hypercube. A detector layer applies a non-linear function like Rectified Linear Unit (ReLU) to learn non-linear representations. Pooling layers make features invariant by reducing the dimensionality of data.

Classification methods based on 1D-CNN (spectral feature-based) and 2D-CNN (spatial feature-based) cannot efficiently classify hyperspectral data as neither utilizes spatial and spectral features together. However, a 3D-CNN can extract spatial-spectral features from the volumetric data. This is due to its ability to incorporate the spectral dimension in addition to the spatial dimensions, which enables it to model and learn more complex spatiotemporal representations. On the flipside, a straightforward generalization of $l \times l$ kernels to their 3-dimensional versions increases the number of operations that must be performed to apply a kernel by a factor of l. Not only the application of these kernels is computationally more costly, but also their training becomes more challenging, as the kernel has more parameters. Taking no other model-changes into account (as, for example in Qiao et al. [52]) we would have to shorten kernel-sides by $\sqrt[3]{l^2}$ to keep computational complexity at roughly the same level.
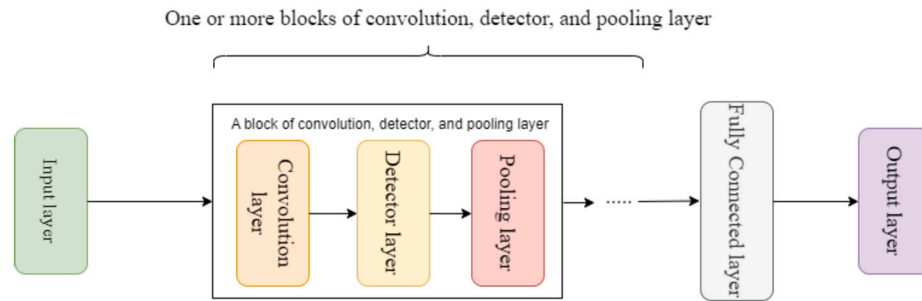
Fig. 2. A basic conceptual CNN architecture. A CNN consists of multiple layers, including convolution, detector, and pooling layers, where each layer serves as the input for the subsequent layer, enabling the extraction of low-dimensional features, learning of non-linear representations, and dimensionality reduction in the network.
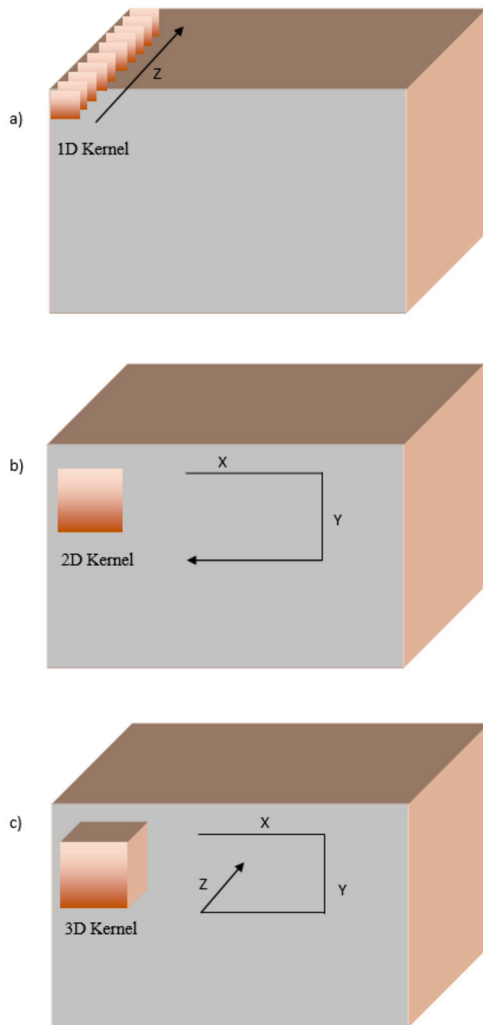

Fig. 3. Movement direction of convolution process using 1D, 2D, and 3D kernels in hypercube. A schematic overview of movement directions for (a) 1D, (b) 2D, and (c) 3D convolutions in CNNs over hypercube is shown. The X and Y directions indicate the movement of the kernel across the spatial dimensions, and Z direction shows the movement across the spectral dimension.

In the following, a description of previous review papers that deal with the topics of CNNs, hyperspectral data, and applications in agriculture and plant research is given. The work of Signoroni et al. [63] is aimed at domain professionals seeking comprehensive insights into the integration of hyperspectral acquisition techniques and deep learning architectures for specific tasks across diverse application domains. This resource caters to machine learning and computer vision experts, of-

fering a nuanced understanding of how deep learning technologies are tailored to effectively process and analyze hyperspectral data, keeping them up-to-date with the latest advancements. In Jiang and Li [29] an examination of how different CNN architectures have been employed in the assessment of plant stress, plant development, and postharvest quality. This review categorizes the studies according to the technical advancements achieved in terms of imaging classification, object detection, and image segmentation. As a result, it highlights cutting-edge solutions for specific phenotyping applications, offering valuable insights into the current state of the field. An overview of state-of-the-art CNN models and visualization techniques for disease diagnosis in plants is given in Joseph et al. [31]. The review given in Gill et al. [21] delves into the plant stress phenotyping, specifically examining the utilization of machine learning and deep learning methodologies. The study encompasses a wide range of high throughput phenotyping platforms, exploring the integration of data from diverse sources. However, it does not discuss 3D-CNN architectures. Finally, a review paper that combines all three topics was provided in Wang et al. [80]. In that work, the authors provide an overview of the application of hyperspectral imaging in agriculture, encompassing areas such as ripeness and component prediction, classification themes, and plant disease detection. Additionally, the study examines recent advancements in hyperspectral image analysis specifically in the context of deep learning models. The review not only highlights the achievements in this field but also outlines the existing challenges associated with deep learning-based hyperspectral image analysis. Moreover, the study presents future prospects and potential directions for further research in this domain.

In this review, we update and complement the paper of Wang et al. [80] with a particular focus on 3D-CNNs and the entirety of the process of creating a high-performant model, including preprocessing of data, band selection, exploration of model architectures, and data visualization.

A deep learning pipeline for HSI classification typically consists of several stages, including data preprocessing, band and feature selection, model design, model training, testing, and evaluation (see Fig. 4). Data preprocessing involves enhancing the quality of raw HSI data, for example, through noise removal, radiometric calibration, and dimension reduction. Feature extraction transforms the raw data into a new space of features, which is expected to be more discriminative for the classification task. Band selection defines a subset of the original spectral bands that is the most relevant for the classification task. Model design is the configuration of hyperparameters of the CNN model, for example, the number and sequence of convolutional layers and dense layers, the activation functions to be used, and the usage of dropout or skip connections. During the training step, the model is optimized by iteratively adjusting its internal parameters (weights and biases) to minimize the error between its predicted output and the actual output on the input data. Finally, the model is evaluated to learn patterns and relationships in the data and to make accurate predictions on new unseen data.
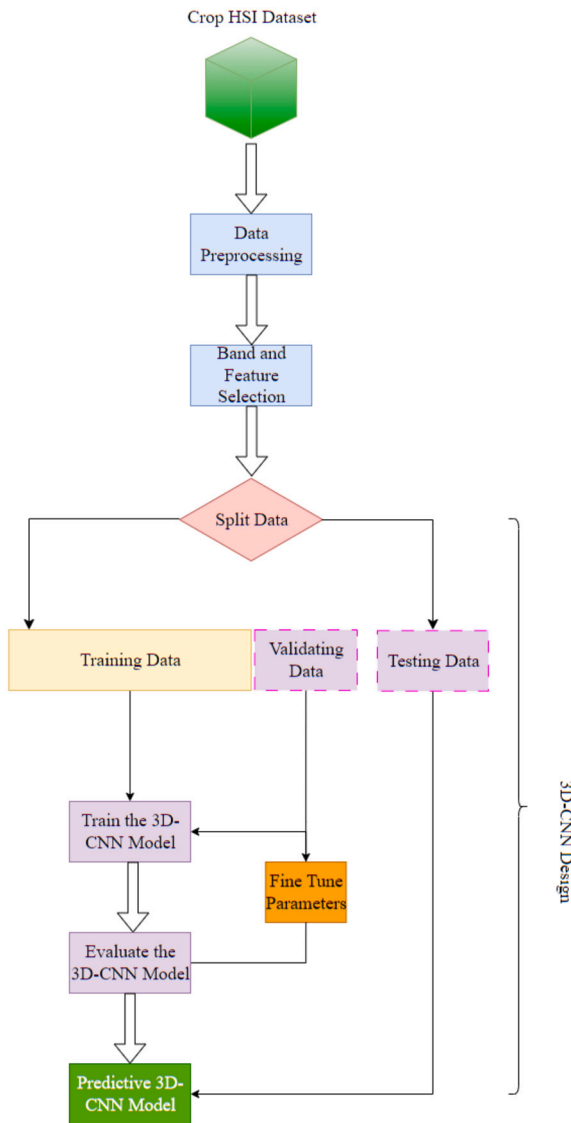
Fig. 4. Typical deep learning pipeline for HSI data classification. First, the discriminative features and bands are extracted from preprocessed HSI dataset of crop. Then, the dataset is split to training, validation, and testing sets. The training dataset is used to train the 3D-CNN model, the validation dataset is employed to assess the model's performance and fine tune its parameters, and the testing dataset serves to evaluate the final performance and generalization ability of the trained 3D-CNN model on unseen data.

4. Data preprocessing

Data preprocessing is a critical step in hyperspectral data analysis, aimed at optimizing the quality and quantity of the data. This step enhances the suitability of the data for downstream tasks such as classification and feature extraction. Preprocessing techniques include patch extraction, radiometric correction and calibration, smoothing, dimension reduction, and background removal. Data augmentation also falls under preprocessing and has the goal of increasing the volume of training data.

4.1. Patch extraction

Patch extraction is a technique that involves dividing an image into smaller images or patches. In the context of HSI, patch extraction has significant advantages for efficient and targeted analysis of specific regions of interest within an image. By extracting image patches that

contain pixels with similar properties, researchers can focus their analysis on areas of the image that are most relevant to diseased or defective areas.

To give a concrete example, Nagasubramanian et al. [45] utilized patch extraction to analyze hyperspectral images of soybean crops. In their study, they extracted spatial patches of size 64×64×240 from an original image of size 500×1600×240, where the first two dimensions define the spatial resolution of the image and 240 denotes the number of spectral bands. By analyzing the properties of pixels within these patches, they were able to extract features that were more representative of the target disease which ultimately improved the accuracy and efficiency of their analysis. Furthermore, patch extraction helps to expand the number of images when there is a lack of data. It also reduces computational time, as processing hyperspectral images of large sizes can be computationally demanding [51,28].

4.2. Data augmentation

Imbalanced and scarce data is a common problem in many machine learning applications (see, for example, [1]) and HSI is no exception. Imbalanced data, a special case of data scarcity, refers to the situation where the number of samples in each class or category of the data is not evenly distributed. This can lead to a bias towards the overrepresented classes in the analysis results, which is particularly problematic if the minority class is of interest.

To address imbalanced HSI data, one common approach is resampling [47], which involves either oversampling the minority class or undersampling the majority class to balance the class distribution. Resampling can be done randomly or using more advanced techniques such as Synthetic Minority Oversampling Technique (SMOTE) [12] or Adaptive Synthetic (ADASYN) sampling [23]. Other widespread data augmentation approaches include transformation techniques such as mirroring [42], rotation [42,13], horizontal and vertical flipping [13,50], and color jittering [50]. Along with above-mentioned methods, patch extraction (Section 4.1) can also be used to address imbalance HSI dataset.

4.3. Radiometric calibration and correction

Radiometric calibration is an essential step in the accurate operation of hyperspectral cameras. It aims to establish a quantitative relationship between the response of the camera sensor (the radiation sensor) and the actual reflectance (radiation level) of an object in a given environment. The calibration process involves assigning a "true" value for either radiation intensity or reflectance to the digital numbers given by the camera that represent recorded outputs for each pixel and spectral channel. To that end, calibrated reflectance standards are utilized, which usually consist of a matte Lambertian reflecting surface to ensure that reflected light is uniform in all directions [60].

The calibration process is then performed as follows: After selecting a radiation source that emits a known type and amount of radiation, the detector is placed in close proximity to the source to measure the radiation. Next, using calibration factors provided by the manufacturer or based on known mathematical formulas (see, for example, [51]) for the particular detector, the expected response of the detector is determined. The measured response of the detector is then aligned with the expected response based on the known radiation level using calibration factors. This calibration process should be repeated periodically with different radiation sources to ensure the continued accuracy and reliability of the detector's measurements.

Even with calibration performed, hyperspectral cameras are susceptible to radiometric errors that can arise from a variety of sources, including sensor drift, electronic interference, light source, and data transmission and recording issues. For example, one of the common radiometric errors which relates to the camera's sensor is stripping. Each sensor consists of multiple individual detectors that sometimes do not function properly due to being out of calibration. Consider a
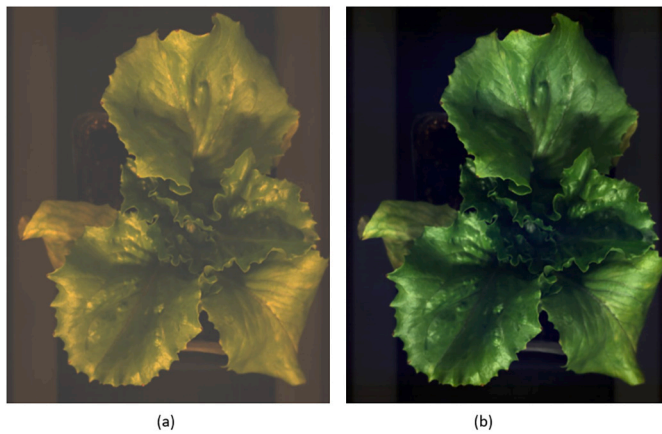
Fig. 5. Black and white calibration of HSI data. (a) The hyperspectral image of lettuce rendered before black and white image correction; (b) The image shows the rendered hyperspectral image after black and white image correction using the Equation (2).

push broom (along-track) camera where its sensor has multiple detectors aligned in a row. When one of them is calibrated slightly different from the adjacent detector the striping effect can occur. In this case, lines predominantly consisting of varying shades of dark and bright pixels formed. Radiometric correction can register and rectify incorrect pixel brightness. To achieve this, a series of procedures are employed including noise correction, de-striping, line-dropout correction [17], and black and white image correction [19,10,13,28]. Fig. 5(b) shows the calibrated hyperspectral image of lettuce using the following computation,

$$\text{Calibrated HSI Data} = \frac{\text{Raw HSI Data} - \text{Dark Ref}}{\text{White Ref} - \text{Dark Ref}}, \qquad (2)$$

where, Raw HSI Data is the image taken by the hyperspectral camera without modification, Dark Ref is the image captured by closing the camera's lens, and the White Ref is the imaged white reference with even and maximum reflectance across the spectral range.

4.4. Smoothing

Smoothing, also known as filtering, is a technique in HSI to improve the quality of the data by reducing noise and artifacts and enhancing the signal-to-noise ratio of the data. This technique in HSI can be broadly classified into two categories: spatial smoothing and spectral smoothing.

Spatial smoothing techniques selectively smooth out certain features in an image by applying filters that amplify or attenuate certain spatial frequencies [17]. Common spatial smoothing techniques include Gaussian filtering, mean filtering [54], median filtering, bilateral filtering [9], and anisotropic diffusion filtering [39].

Spectral smoothing techniques operate on the spectral domain of an HSI image by smoothing the intensity values of neighboring spectral bands. As in spatial smoothing the goal is to remove noise [77], reduce artifacts, and enhance features in the image. Common spectral smoothing techniques include moving average filtering [77], Savitzky-Golay (SG) filtering [78], Fourier filtering, wavelet filtering, and principal component analysis (PCA). For example, Cao et al. [10] eliminated the random noise that was present in the spectral data of the different region of interests (ROIs) using SG filtering. The filtering process makes it easier to identify the true signal of the sample and remove interference caused by size and structure differences between the ROIs. Similarly, Jung et al. [33] and Jia et al. [28] used a SG filter to smooth out the spectral data and reduce the effect of noise.

As Vaiphasa [77] points out, care must be taken when applying smoothing techniques. The subjective selection of smoothing filters in

hyperspectral remote sensing studies can negatively impact the statistical properties of the spectral data, which can, in turn, affect subsequent analyses. To preserve the statistical properties of the HSI data, the selection of smoothing filters should be done through a comparative t-test method that identifies the filter with the least statistical disturbances. By using this approach, it is possible to mitigate the negative effects of smoothing filters and ensure the reliability of subsequent analyses based on statistical class models.

4.5. Dimension reduction

The large number of spectral bands within hyperspectral data often makes it challenging to process and analyze HSI data. Dimensionality reduction techniques retain relevant information, while allowing the model to work on smaller hypercubes downstream. These techniques are split into linear and non-linear ones.

4.5.1. Linear techniques

Linear dimension reduction in HSI refers to a set of statistical and machine learning techniques that aim to find a lower-dimensional linear manifold in the high-dimensional space that captures the essential spectral information of the data [35]. By embedding the data into this lower-dimensional linear space, the dimensionality of the data is reduced while preserving as much of the spectral information as possible. PCA and Random Forest (RF) are two examples of this technique, and demonstrations of both can be seen in Cao et al. [10].

The process of PCA involves first zero-centering the input spectral matrix (i.e., the matrix of the spectral bands) and computing its covariance matrix. Next, the eigenvectors and eigenvalues of the covariance matrix are calculated, and the eigenvalues are sorted in descending order. The top k eigenvalues are then selected and the corresponding eigenvectors space computed. The k-dimensional data is obtained by projecting the original spectral matrix into the new space using the selected eigenvectors.

On the other hand, an RF evaluates the importance of each wavelength by randomly replacing it and measuring its effect on the accuracy of the thus trained CNN model. To accomplish this, the RF builds many decision trees, and each tree is trained on a different subset of the data. The algorithm calculates the importance score for each wavelength by measuring the change in the prediction error rate before and after randomly replacing the wavelength in the out-of-bag data. The wavelength with the highest importance score is selected, and this process is repeated until the desired number of wavelengths is achieved. The dimensionality of the hyperspectral data is thus reduced, while retaining the most important information for accurate predictions.

Further linear algorithms for dimension reduction of HSI data are described in Fırat et al. [18] and encompass Independent Component Analysis (ICA) and PCA-based algorithms like Incremental PCA (IPCA), Sparse PCA (SPCA), and Randomized PCA (RPCA).

4.5.2. Non-linear techniques

Non-linear techniques can handle data with complex and non-linear structures. Yang et al. [83] classified these techniques into Kernel-based and manifold learning [83]. Kernel-based techniques, like Kernel PCA (KPCA), use non-linear mappings to transform data into higher-dimensional feature spaces first on which linear techniques can be applied. Manifold learning algorithms, on the other hand, aim to directly discover the intrinsic non-linear structure of data. See Khodr and Younes [35] for well-known manifold learning techniques, including Isometric Feature Mapping (Isomap), Locally Linear Embedding (LLE), Local Tangent Space Alignment (LTSA), Diffusion Maps, Sammon's Mapping (SM), and Locality Preserving Projections (LPP).

4.6. Background removal

The data captured in hyperspectral images contains both foreground, called the ROI, and background objects. In scenarios where

the target object does not cover the entire scanning area the signals from background objects can interfere with the data analysis, i.e., the background can contain noise that needs to be filtered out [78]. This is especially true when dealing with images that exhibit color gradients. By masking the background from the data, researchers can focus on the spectral signature of the ROI. This leads to improved target detection in classification task. It also reduces the computational complexity of subsequent processing steps including the training 3D-CNN models [51].

There are various traditional techniques to extract he ROI from the hyperspectral image. These techniques can be classified into several categories based on their underlying principles and are discussed separately below.

4.6.1. Spectral similarity-based methods

These methods work based on the similarity between the spectra of the pixels within an image. Examples of such methods include Spectral Angle Mapper (SAM) [37] and Spectral Information Divergence (SID) [53]. SAM computes the spectral angle

$$\alpha = \cos^{-1}\left(\frac{\sum_{i=1}^{\lambda} R_i \cdot T_i}{(\sum_{i=1}^{\lambda} R_i^2)^{\frac{1}{2}} \cdot (\sum_{i=1}^{\lambda} T_i^2)^{\frac{1}{2}}}\right) \tag{3}$$

between the target reference spectrum R and each pixel spectrum T for all spectral bands λ in the hyperspectral image. This results in a similarity measure that is insensitive to illumination variations [2]. On the other hand, SID works based on the concept of information theory (entropy). It compares the spectral information content of each pixel to a reference spectral information content. In general, SAM is better suited for well-defined spectral variations and low background noise, while SID is more robust to complex background noise and illumination variations.

4.6.2. Statistical-based methods

These methods leverage statistical techniques to identify ROIs that share similar spectral properties and produce a set of uncorrelated components capturing different aspects of spectral variability within the image. For example, Minimum Noise Fraction (MNF) [43] is specifically designed to reduce the impact of noise in the data by separating the noise and signal components of the HSI data. This makes it particularly useful for operating on noisy HSI data but at the expense of more computational time. Another technique, ICA (see Section 4.5.1), aims to separate the mixed signals into their independent components, providing a more flexible approach to identify subtle spectral differences between ROIs. Finally, PCA [13] decomposes the original HSI data into orthogonal components that represent the directions of maximum variance in the data. In general, PCA is computationally more efficient than ICA and MNF, as it involves a simpler mathematical transformation that uses standard matrix operations.

4.6.3. Spatial-based methods

These methods exploit the spatial correlation present in the image to differentiate between pixels within and outside the ROI. Typically, these methods apply morphological operations or spatial filtering techniques to extract features such as edges or texture, which are then used to segment the image into ROIs. The Morphological Attribute Profile (MAP) [14] and the Spatial-Spectral Endmember Extraction (SSEE) algorithm [49] are examples of spatial-based methods that use mathematical morphology and spatial filtering, respectively.

SSEE first identifies the endmembers, or pure spectral signatures, present in the HSI data, and then uses a spatial clustering algorithm to group adjacent pixels with similar spectral properties into ROIs. On the other hand, MAP applies a series of morphological opening and closing operations to the image to identify connected regions of pixels with similar morphological attributes, such as size and shape. These connected regions can then be used as ROIs. These methods are computationally efficient and can be useful in scenarios where spectral information alone

is not sufficient for accurate ROI extraction, such as in cases of low spectral contrast or high noise levels.

4.6.4. Hybrid methods

These methods combine techniques to improve the accuracy and efficiency of the ROI extraction process. One approach can be combining statistical-based techniques such as PCA, ICA, or MNF with spatial-based techniques such as MAP or SSEE. For example, by combining MAP and PCA we can exploit both the spatial and spectral information [70] in the HSI data for ROI extraction. These hybrid methods can be effective in cases where neither spatial nor spectral methods alone are sufficient for accurate ROI extraction.

4.6.5. Machine learning-based methods

Machine learning algorithms can also be used to extract the ROI. This process involves training a model using labeled data to identify and extract regions with specific spectral characteristics. Once trained, the model can be used to predict the presence of those characteristics in unlabeled data and extract ROIs. This procedure is computationally efficient and allows for the extraction of subtle and complex patterns that may not be easily identifiable through traditional methods. Some instances of such techniques are Support Vector Machines (SVM) [6], RF [7], and CNN [40,79].

4.6.6. Software-assisted manual annotation

Manual definition of the ROI can be assisted by software specifically built to handle HSI data. Amongst these are for example ENVI [38] and Spectronon [55] as commercial products, as well as the MATLAB Hyperspectral toolbox. Examples for open-source and free software are SeaDAS [46], the Orfeo ToolBox [11], and RSGISLib [8]. These packages provide HSI data analysis for a wide range of tasks discussed before. An example of software-assisted annotation is the work of Jin et al. [30] in which the authors generated an ROI using ENVI by manually selecting the tissues or areas of interest in the false color images of HSI data.

Researchers can utilize such software in addition to the above discussed algorithms. For example, Gao et al. [19] developed an open-source software which is specifically designed for analyzing HSI data of seeds. It is able to remove the background of the HSI data and produces a binary mask using a user-defined minimum and maximum intensity threshold along with a component-searching algorithm. The results are an accurate segmentation of the seeds even when they overlap.

5. Band and feature selection

The selection of spectral bands in HSI is another crucial preprocessing step for classification tasks. It entails the identification of a subset of the most discriminative spectral bands from all available bands. This process is instrumental in reducing data dimensionality by eliminating redundant bands, thereby significantly reducing the computational overheads of downsteam tasks. Furthermore, the careful selection of spectral bands can also reduce the effects of noise in the data. In general, based on the survey of Sun and Du [71], band selection mechanisms can be categorized into six groups as follows (another classification of methods is presented in Sawant and Prabukumar [58]).

5.1. Ranking-based selection

Ranking-based band selection methods evaluate the significance of each spectral band based on a predetermined criterion and choose the most important bands in a sorted order. These methods can be categorized into two types: supervised and unsupervised. In supervised ranking-based methods, labeled training samples are utilized to determine the importance of each spectral band, while in unsupervised ranking-based methods, statistical properties of the data are used for the same purpose.

One example of such method is spectral differentiation. Qi et al. [51] employed first and second order differentiation to decrease the computational complexity of HSI data containing 204 bands. The first derivative can effectively pinpoint areas in the spectrum where the rate of change is highest. This indicates the presence of sharp spectral features such as absorption or emission lines. Moreover, it allows for the selection of bands that capture these features and thus, provide critical information for classification or detection tasks. The second derivative is useful for identifying regions of the spectrum where the rate of change of the first derivative is highest, signifying the presence of spectral curvature. The selection of bands that capture the shape of the spectral signature, based on this information, can enhance the differentiating power in classification or detection tasks. Likewise, Jung et al. [33] achieved an increase in accuracy using spectral differentiation and expansion of the input in the vertical direction of the raw data. This technique was applied in addition to SG smoothing (see Section 4.4) to further improve data quality.

5.2. Searching-based selection

Searching-based band selection methods involve the creation of a criterion function such as Euclidean distance and Bhattacharyya distance [26] to evaluate the performance of each spectral band based on a specified optimization objective. The first step involves creating an initial subset of bands, followed by an assessment of the criterion function for the subset. The next step is applying a searching strategy to identify the best subset of bands that maximizes the criterion function, and evaluating the selected subset based on data classification performance. This iterative process continues until the desired level of performance is reached. Searching-based methods largely depend on the quality of the criterion function and the optimization strategy employed. Incremental searching [81,57], updated searching [20,62], and eliminating searching [68] are among the commonly utilized strategies.

5.3. Clustering-based selection

Clustering-based methods for hyperspectral band selection group bands into clusters and select representative bands from each cluster to create a final subset. These algorithms can be unsupervised [27,82], supervised [44] or semisupervised [66,65]. The selection of representative bands is typically performed using information measurements, such as mutual information or Kullback–Leibler divergence. Commonly used clustering techniques are K-means, affinity propagation, and graph clustering. K-means selects the best cluster centers that minimize the sum of distances to a set of putative center candidates. Affinity propagation selects exemplars by considering the correlation or similarity among bands and the discriminative capability of each band.

5.4. Sparsity-based selection

Sparsity-based techniques for band selection rely on sparse representation or regression to identify representative bands. The most common of these methods are discussed separately below.

5.4.1. Sparse nonnegative matrix factorization-based methods

These methods [41] break down the hypercube into a set of building blocks, which are both nonnegative and sparsely encoded. This promotes a feature extraction process that combines these building blocks to create a parts-based representation of the original data. The goal of this method is to identify the most informative bands of the HSI data matrix by optimizing an objective function that includes sparsity constraints.

5.4.2. Sparse representation-based methods

Sparse representation-based methods [86,72] use pre-defined or learned dictionaries to select informative bands of the HSI data matrix based on their sparse coefficients. These methods rank the bands according to the frequency of their occurrence in the sparse coefficient histograms. In some cases, sparse representation-based methods can also be designed to solve multiple tasks simultaneously, and an immune clonal strategy can be used to search for the best combinations of informative bands.

5.4.3. Sparse regression-based methods

Sparse regression-based techniques [69,15] transform the band selection problem into a regression problem and estimate the most representative bands by solving a sparse regression problem. These methods can also include sparsity constraints to encourage the selection of only the most informative bands for the regression model.

5.5. Embedding-learning based selection

Embedding-learning based methods aim to learn a low-dimensional representation of the spectral data, also known as an embedding, that captures the most salient features of the data. There are several types of embedding learning-based methods that can be used for band selection, including autoencoders [75], Deep Neural Networks (DNN) [87], and CNNs [61]. Autoencoders learn a compact representation of the input data by training a neural network to encode the input into a lower-dimensional space and then decode it back into the original space. DNN, on the other hand, consists of multiple layers of interconnected neurons. This architecture enables the network to learn complex patterns and features in a hierarchical manner, empowering it to extract high-level representations from the input data. Meanwhile, CNNs can learn spatially invariant features from image data.

These techniques aim to learn a set of parameters that minimize an objective function that measures the model's performance on a particular task, such as classification or target detection. Band selection is integrated into the optimization process by constraining the learning algorithm to focus on a subset of the available bands, or by assigning weights to each band that reflect its relevance to the task at hand. The resulting model can then be utilized to predict the class label of new samples or to detect the presence of specific targets in the image. For example, Chen et al. [13] employed a DNN based binary classification to identify the foreground and background regions of the HSI data. Subsequently, connected component labeling algorithms and edge contours were used to isolate the ROI from the image for further analysis.

Moreover, Jia et al. [28] implemented a CNN-based band selection module that works based on a group convolution technique, which involves applying a 1×1 one-dimensional convolution (equivalent to a scalar multiplication) to each band of the input hyperspectral image independently. This technique helps to overcome the problem of mutual interference between different channels. The weights of the convolutional kernel are updated in the early stage of the network training using a loss function and an auxiliary classifier (see also the next Section 6). The weights represent the importance of each band, with a higher absolute value of weight indicating greater importance of the corresponding band.

5.6. Hybrid-scheme based selection

Hybrid-scheme based methods involve combining multiple band selection techniques to select the most appropriate bands. A popular combination is clustering and ranking [84,16], where clustering is used to group bands and ranking is used to select the most important bands within each cluster. Other hybrid methods combine clustering with searching or combine ranking with searching to further optimize band selection.

6. Network architecture design: feature extraction and classification

The objective of neural network architecture design is to create a model that can effectively learn from the input data and generalize well

Table 1
Non-hybrid 3D-CNN-based architectures for detection of diseased and defected hyperspectral images of crop.

CNN Model	Information	Reference
3D-CNN-based	**Dataset**: 111 hyperspectral images of soybean **Type of Disease**: Charcoal rot **Imaging Device**: Pika XC hyperspectral line imaging scanner **Spectral range**: 400–1000 nm **GPU**: NVIDIA Tesla P40	Nagasubramanian et al. [45]
3D-CNN based on AlexNet	**Dataset**: 40 hyperspectral images of Grapevine groups **Type of disease**: GVCV **Imaging Device**: SPECIM IQ **Spectral range**: 400–1000 nm **GPU**: -	Nguyen et al. [47]
HyperSeed	**Dataset**: 200 rice seeds (274,641 pixels) **Type of defect**: Heat stress **Imaging Device**:Micro-Hyperspec Imaging Sensors, Extended VNIR version **Spectral range**: 600-1700 nm **GPU**: -	Gao et al. [19] GitHub Repository
3D-CNN-based	**Dataset**: Above 200 strawberry leaves (3,110 ROIs) **Type of disease**: Gray mold **Imaging Device**: Corning microHSI **Spectral range**: 400-1000nm **GPU**: NVIDIA RTX3090 X (24 GB memory)	Jung et al. [33]

to new, unseen data. In order to achieve this for HSI classification, a network architecture must be developed that can capture the complex spectral and spatial information present in the data.

As an integral aspect of designing a neural network, the decision of the number of layers to use is a crucial one. While the inclusion of more layers in a network has been shown [76,32] to improve performance, it can simultaneously present several challenges. A notable challenge arises from the potential occurrence of vanishing or exploding gradients [73], where the gradient signal becomes too small or too large as it backpropagates through the layers during training (see also Subsection 6.2.1). Such a phenomenon makes it difficult for the network to learn and adjust its weights effectively.

Moreover, as the number of layers and parameters increases, the risk of overfitting rises, which is characterized by the network's ability to perform exceptionally well on the training data, but not generalize well on unseen data. Consequently, the network may become computationally expensive to train and use due to its high processing power and memory requirements. Additionally, gradient computation time increases, thereby impeding the training's efficiency. Hence, careful consideration of these challenges and trade-offs is imperative to designing a network that balances complexity and performance.

For this review, we classified 3D-CNNs into hybrid and non-hybrid structures. A network is hybrid if it includes either specific module(s) that improve feature extraction, accuracy, and performance or if it integrates 2D-CNN within the 3D-CNN architecture. In the following, we present 3D-CNN models for the classification of diseased and defected crop using HSI data. A summary of these architectures is also given in Tables 1 and 2.

6.1. Non-hybrid networks

Nagasubramanian et al. [45] presented a 3D-CNN model to classify healthy and diseased crop. This model consists of two convolution layers with max pooling layers, and two Fully Connected (FC) layers, trained using the Adam optimizer. To prevent overfitting, dropout mechanisms were used after the first max pooling and first FC layer. A Weighted Binary Cross Entropy (WBCE) function of the form

$$L_{WBCE}(y, \hat{y}) = -[\beta \cdot y \log(\hat{y}) - (1 - y) \log(1 - \hat{y})] \tag{4}$$

was implemented to address imbalanced training data. Here y and $\hat{y}$ represent binary variables for whether the ground truth and predicted

result belong to a given class, respectively (see also Section 4.2 for a discussion on how to counteract imbalances in datasets). Using the coefficient β the WBCE loss function assigns higher weights to the minority class, for example, the false negatives rate decreases if β is set higher than 1, while setting it smaller than 1 reduces the false positive rate.

Although the above technique can classify and detect diseased crop, detection of asymptomatic diseased crops at early stage and differentiating it from the healthy crops can be more challenging. In this respect, Jung et al. [33] developed a 3D-CNN model that improves classification accuracy of the asymptomatic diseased crop without modification and preprocessing of input HSI data. The model consists of four 3D convolution layers in which the first and fourth layers each are followed by a 3D max pooling layer and a batch normalization. The rest of the convolution layers are only followed by batch normalization. The output of the last 3D convolution layer is passed through a global average pooling and two dense layers. In Jung et al. [33] the results were improved further by preprocessing the input HSI data which went through spectral differentiation, vertical expansion, and smoothing.

Automating the classification procedure of HSI data using software can effectively mitigate the time-consuming process of implementing a deep learning pipeline. In this respect, Gao et al. [19] developed an open-source software that classifies seeds at pixel level using 3D-CNN. Though at first it was developed for a specific crop (rice), the test experiments over other seeds are promising. This software utilizes a 3D-CNN that consists of two 3D convolution layers, with two and four 3D convolution kernels for the first and second layers, respectively. The output is then flattened via one FC layer and classification is performed using a softmax function. However, this software as part of its feature extraction process does not consider the global feature relationship which can improve its ability to recognize and classify seeds accurately.

Nguyen et al. [47] implemented a 3D-CNN for classification of a small size dataset. The feature extraction part is implemented based on AlexNet [36] and constitutes 5 convolutional layers followed by a flattening layer. The input layer takes an input of size 512×512×203, where 203 stands for number of bands. The first two convolutional layers are followed by a max pooling layer and a batch normalization. After that each convolutional layer is followed by a max pooling layer and after the last max pooling layer batch normalization and flattenings is applied before feeding the result into a RF or SVM for binary classification (healthy or diseased crop).

Table 2
Hybrid 3D-CNN-based architectures for detection of diseased and defected hyperspectral images of crop.

CNN Model	Information	Reference
Hypernet-PRMF	**Dataset**: 16 hyperspectral images of peanut **Type of disease**: Mold **Imaging Device**: SOC710E portable hyperspectral imager **Spectral range**: 400–1000 nm **GPU**: NVIDIA Tesla P100 GPU (12G)	Liu et al. [42]
Deep ResNet 3D-CNN	**Dataset**: 16,346 hyperspectral images of blueberry **Type of defect**: Distinguishing decayed and sound blueberries **Imaging Device**: - **Spectral range**: 400-1000nm **GPU**: -	Qiao et al. [52]
SDC-3DCNN	**Dataset**: Rice leaves (Number of taken samples are not determined.) **Type of disease**: BLB **Imaging Device**: Raptor EM285 **Spectral range**: 378.28–1033.05 nm **GPU**: NVIDIA GeForce RTX 2080Ti GPU and the AMD Ryzen 5-1600 Six-Core processor @ 3.20 GHZ CPUs	Cao et al. [10]
2D-3D-CNN (Defect detection module of RT-CBDIA)	**Dataset**: 1026 coffee beans **Type of defect**: Black, insect-damaged, and shell **Imaging Device**: Imec XIMEA snapshot sensor **Spectral range**: 660-980 nm **GPU**: GPU of GEFORCE GTX1660 Ti and a RAM of 16 GB	Chen et al. [13]
ResNet	**Dataset**: 210 lemons **Type of defect**: Bruise **Imaging Device**: It was not determined, however was provisioned by Noor Imen Tajhiz Co.. **Spectral range**: 400-1100 nm **GPU**: Trained on Google Colab	Pourdarbani et al. [50]
PLB-2D-3D-A	**Dataset**: 15,360 potato leaves **Type of disease**: PLB **Imaging Device**: Specim IQ **Spectral range**: 400-1000 nm **GPU**: NVIDIA Tesla V100	Qi et al. [51]
Y-Net	**Dataset**: 200 diseased corn leaves (extracted 6,264 regions) **Type of disease**: Brown spot and anthracnose **Imaging Device**: An HSI system provided by Head Wall **Spectral range**: 400-1000 nm **GPU**: RTX 3090 24 Gb	Jia et al. [28]

6.2. Hybrid networks

6.2.1. 3D-CNN architectures based on ResNet

To address the challenges of vanishing or exploding gradients Qiao et al. [52] proposed to leverage residual convolutional blocks within a 3D deep ResNet architecture. This approach reduces the number of channels through the use of identity residual blocks and a convolutional residual blocks. The identity residual block maintains the same input and output dimensions, while the convolutional residual block changes the number of channels. The use of a l×l×l convolution kernel as the shortcut in the convolutional residual block reduces the number of parameters and computational complexity.

To further improve efficiency, Qiao et al. [52] adopt a bottleneck structure, that reduces the number of required convolution operations. Each convolutional layer is followed by a batch normalization layer to prevent vanishing gradient and enhance convergence rate. The network uses Exponential Linear Unit (ELU) as the non-linear activation function, which addresses the problem of dying neurons in ReLU.

In order to identify the appropriate hyperparameters, the authors employed a Tree-structured Parzen Estimator (TPE) as an optimization algorithm. The TPE utilizes a probabilistic model to approximate the distribution of the objective function and guides the search for optimal hyperparameters. For the classification, a 7×7×1 global pooling layer and a FC are utilized. This model halves the number of parameters and improves the computational time up to 10%. Moreover, the study of Pourdarbani et al. [50] over performance of well-known architectures in detection of defective crop demonstrates that residual connections achieve higher accuracy, while training faster despite having more parameters.

6.2.2. Hypernet-PRMF network

Liu et al. [42] presented a feature pre-extraction and a multi-feature fusion block to extract peanut characteristics from hyperspectral data. The feature pre-extraction includes constructing a Peanut Recognition Index (PRI) based on two informative bands to distinguish healthy, moldy, and damaged peanuts.

The multi-feature fusion block is a technique used in image segmentation to fully extract spatial and spectral features from HSI data. This technique involves using multiple types of convolution kernels including 2D convolution for common texture features, separable convolution for increased feature diversity, depthwise convolution for band feature extraction, and 3D convolution for spectral change information. The convolutions are concatenated after normalization and activation functions to enhance diversity and improve recognition accuracy.

Moreover, the authors employed feature pre-extraction and multi-feature fusion block techniques in their proposed peanut recognition model, called Hypernet-PRMF network. This model works at both peanut- and pixel-level recognition. The model consists of four parts: feature pre-extraction, down-sampling, up-sampling, and prediction. The feature pre-extraction part enhances the differentiation between different peanut features. The down-sampling part reduces the size of the image while increasing the number of convolution kernels, whereas the up-sampling part reconstructs the image while reducing the number of convolution kernels. The prediction works based on the softmax

function and the class of maximum predicted probability is chosen as the final recognition result. The model achieves pixel-wise recognition accuracy with the use of the watershed segmentation algorithm. This technique has the potential to be employed for detection of other crops like sorghum.

6.2.3. Spectral dilated convolution 3D-CNN

Cao et al. [10] proposed a spectral dilated convolution (SDC)-3D-CNN model to detect crop's asymptomatic diseases at an early stage. This model consists of SDC modules along with residual blocks that prevent the gradient vanishing problem. SDC extends the idea of dilated convolution which expands the receptive field of convolution kernels without augmenting the model's parameterization. Receptive field is the portion of the input space needed to create a filter at any convolutional layer. The 3D-SDC extends the receptive field of convolutional kernels to the spectral dimension. It works based on the principle of applying a filter to an input with intermittent intervals, which are dictated by the spectral dilation rate.

The network was tested with top 50, 100, 150, and 200 significant wavelengths extracted by RF and Principal Components (PCs) of the same ranking by PCA along with different spectral dilation rate to detect healthy, asymptomatic, and symptomatic crop. The experiment result shows higher detection performance of the network using top the 50 important features extracted by RF at a dilation rate of 5.

6.2.4. Merged 2D- and 3D-CNN architectures

Chen et al. [13] developed a 2D-3D-CNN for real-time crop defect detection. This network is the detection module of a real-time coffee-bean defect inspection algorithm (RT-CBDIA). The network consists of a 2D-CNN and a 3D-CNN, the former one is responsible to extract spatial features and the latter one is accountable for extraction of spectral features. Combining these two networks can boost feature extraction by providing robust and discriminative spectral-spatial features.

The 3D-CNN is comprised of two convolution blocks with the same structure as in the 2D-CNN except that 3D convolutions and pooling layers are used. The two networks run simultaneously and their last pooling layers will be merged and fed to a FC layer and then a dropout layer to avoid overfitting. Finally, a softmax layer determines each crop health status.

Likewise, Qi et al. [51] fully extracted spatial-spectral features by merging 2D- and 3D-CNN architectures using AttentionBlock [85] and Squeeze-and-Excitation (SE)-ResNet [25]. To accomplish this, the model first creates a neighborhood block of size 11×11×10 around a center pixel from the input image. Then, 2D convolution operations are used to extract spatial correlation features from the neighborhood block, and 3D convolution operations are used to capture spectral correlation features. The model uses four 2D convolutional layers and four 3D convolutional layers to capture feature maps of various spatial and spectral dimensions. By using different sizes of convolution kernels and downsampling steps, varied types of information can be captured.

Finally, the extracted feature maps are fused together to create a final set of feature maps that contain valuable and pertinent information required for effective classification. Herein, AttentionBlock and SE-ResNet play important role, as outlined immediately below.

An AttentionBlock is used to highlight important information in the fused spectral space feature map. It works by considering the similarity between each pixel in the feature map and weighting the relevant pixels with higher importance. This is achieved through a series of 2D convolutions with a kernel size of 1×1 to transform each pixel into an λ-dimensional vector, where λ represents the number of feature channels in the input tensor. The similarity between any two pixels is then calculated using the dot-product of their transformed vectors, and the results are weighted using a softmax function.

The output of an AttentionBlockt is a feature map that emphasizes the relevant information while suppressing irrelevant information. This process allows AttentionBlocks to focus on the relevance between pix-els in the entire feature map, rather than just the spatial range of the convolution kernel size used in traditional convolution and pooling operations. This results in better classification results with little computational complexity.

In order to enhance the representational power of CNNs, SE modules as a type of attention mechanism can adaptively recalibrate the feature maps. It does so by capturing channel-wise feature dependencies through a squeeze operation, followed by an excitation operation that learns how to weight the importance of each feature map. This mechanism allows the model to pay more attention to salient channel features and disregards the less significant ones. By integrating AttentionBlocks and SE-ResNet, the network of Qi et al. [51] can better generalize in classification and achieve higher accuracy.

Moreover, in the context of detection of two similar crop diseases that are indistinguishable to the naked eyes, a recent study by Jia et al. [28] developed a new network called Y-Net. The Y-Net model takes in 10×10×203 hyperspectral data cubes as input and it consists of a channel attention mechanism, a band selection module with auxiliary classifier, a 3D-2D-CNN architecture, and a classification module.

A CNN architecture is employed to conduct the band selection, where 1×1 1D convolutions are assembled to modify the parameters of the convolutional kernel in the early phase of the network training. The magnitude of the weight of the convolution kernel is indicative of the relevance of the band, with greater absolute weight values signifying more distinctive bands. The use of group convolution helps in preventing any hindrance from nearby bands, while the ReLU activation function is adopted for fast network convergence without the problem of saturation. The output of this step will be given to an auxiliary classifier that enables early-stage weight updating in the band selection block.

The auxiliary classifier module updates the loss function of the Y-Net model:

$$L(y, \hat{y}) = (1 - \theta) \cdot L_{\text{Final Classifier}}(y, \hat{y}) +$$
$$\theta \cdot L_{\text{Auxiliary Classifier}}(y, \hat{y}) + \beta \cdot \sum_{j=1}^{n} W_j, \tag{5}$$

where y is the ground truth label, $\hat{y}$ is the predicted label, and θ and β are hyperparameters that control the trade-off between the two losses and the sparsity of the band selection module, respectively. W_j is the weight of the jth band in the band selection module and n is the number of total bands. The loss is a combination of the cross entropy losses of the final classifier and the auxiliary classifier and the sum of the weights of the band selection module. The purpose of this combination is to control the classification accuracy while updating the weights of the band selection layer. The adjustment factor θ gradually decreases as the number of training iterations increases. The presence of this adjustment factor enables the Y-Net model to update the weights in the band selection module in the early stages of training and gradually shift towards training the final classifier to learn a more accurate classification model. Additionally, the auxiliary classifier helps to constrain the weight sparsity of the band selection module, ensuring that the score of unimportant features is close to zero.

The results of Jia et al. [28] show that by removing nonessential and nondiscriminative bands the accuracy of the Y-Net model increases and reduces the model size and the number of parameters. Moreover, since the band selection module is integrated into an overall architecture, the training time does not significantly increase.

7. Visualization techniques for HSI classification decisions

Visualization techniques can be employed to observe the contribution of pixels in the classification decision. These techniques allow us to identify the pixel locations associated with the most important spectral bands that play a crucial role in the final classification results.

Saliency maps [64] are one of the essential and traditional visualization tools to identify the most sensitive regions (crucial pixels) in an image with respect to a model's predictions. This technique works by computing the gradient of the output class score with respect to the input image. This gradient represents how much each pixel in the input image contributes to the final classification decision. Next, the absolute values of these gradients are summed across the channels to obtain a saliency map, which highlights the most salient regions of the input image for the predicted class.

For example, Nagasubramanian et al. [45] discovered that saliency maps can help to locate the most sensitive pixel locations in infected crop images, which are often the severely infected areas. Conversely, both healthy and infected crop images had saliency map gradients that were primarily focused around the mid-region of the crop stem, highlighting the stem's importance in crop classification. Moreover, Cao et al. [10] observed that the significant wavelengths extracted by RF from raw HSI data overlaps the saliency-sensitive wavelengths. More importantly, saliency maps can determine significant wavelengths for classification which are not extracted by RF.

Another visual explanation technique for CNN decision is the Gradient-weighted Class Activation Mapping (Grad-CAM) [59]. It was introduced as an improvement over CAM [88]. CAM involves modifying a pre-existing CNN model by replacing the final FC layer with a global average pooling layer, which retains essential channel information while reducing the spatial dimensions. This modification enables the utilization of feature maps from the preceding layer. By applying learned weights to these feature maps through global average pooling, CAM generates a map that highlights the crucial regions associated with the predicted category. However, CAM provides a coarse localization of the important regions within an image. It highlights the regions that contribute most to the predicted class, but it does not provide precise boundaries of those regions. To address this limitation, Grad-CAM was introduced as an extension to CAM.

Grad-CAM enhances the CAM approach by incorporating gradient information. Similar to CAM, Grad-CAM also utilizes a global average pooling layer after the last convolutional layer to obtain importance scores for each channel in the feature maps. However, instead of learning separate linear models for each class, Grad-CAM calculates the gradients of the predicted class score with respect to the feature maps. These gradients are used to weigh the feature maps and enables Grad-CAM to identify more intricate features that play a significant role in the classification decision.

In addition, there are further statistical techniques that help in analyzing the importance of individual features (spectral bands) of HSI data in a classification. Light Gradient Boosting Machine (LightGBM) [34,19] is one of such methods that uses decision trees in calculating feature importance. It builds decision trees in a leaf-wise manner, meaning that the algorithm grows the tree by adding new leaves one at a time which is computationally faster compared to level-wise approach by selecting the best split based on the maximum reduction in loss function for all leaves in the tree. Therefore, LightGBM calculates the importance of each feature by evaluating how much it contributes to the reduction in loss function across all trees. The feature importance score is calculated by summing up the number of times a feature is used to split the data across all trees, weighted by the improvement in accuracy achieved by each split. Features that are used more frequently and result in larger improvements in accuracy are assigned higher importance scores.

8. Discussion and conclusion

In this study, we conducted a comprehensive review of 3D-CNN-based models applied in the domain of agriculture using non-UAV-based HSI data. Our analysis delved into diseased and defective crops, focusing on the structures and efficiencies of the models, the quantity of datasets utilized, and the necessary preprocessing steps. This review indicates the advantages of 3D-CNNs in capturing spatial-spectral in-

formation within hyperspectral data, enabling them to outperform 1D- and 2D-CNN in hyperspectral image classification. With the aim of assisting computer vision experts and agriculture-domain researchers in tackling HSI classification tasks for crops experiencing stress, this comprehensive review provides valuable insights and guidance.

In general, HSI holds great potential for detecting subtle changes in crop growth and development, making it a promising technique for diagnosing crop diseases and defects. Despite this potential, our study indicates that there is still limited research conducted that use 3D-CNNs in this context. Furthermore, the studies that are performed are often very application-specific and it is unknown how well the performed methods and models generalize to a broader range of applications. We identify three major challenges that must be overcome to achieve a broader adoption of HSI in general and the usage of 3D-CNN for HSI classification problems: limited availability of hyperspectral data, computational complexity of 3D-CNN models, and the costs of hyperspectral imaging hardware. In the following we address each of these challenges in more detail and offer ideas on how to overcome or avoid them.

Limited availability of varied hyperspectral data on diseased and defective crops presents a significant challenge for researchers, farmers, and stakeholders in the agriculture industry who rely on data to make informed decisions. The lack of data in this area of the research hampers efforts to understand the extent of the problem and develop effective solutions. It makes it difficult to track the progress and success of any initiatives aimed at improving crop health and reducing the prevalence of disease and defects. To address this issue, there is a need for increased investment in data collection and data sharing, as it has been done in the past, for example, for RGB-data (e.g., Beck et al. [4,3]). Large-scale HSI data collection and publicly available datasets will allow research groups to develop the next generation of models, even if they have no access to the otherwise required hardware and plant material. Even with small datasets there are also techniques to models beyond the specific application case they had been trained on. In recent years, transfer learning and active learning techniques have been increasingly used together to tackle the challenges posed by limited data. By leveraging knowledge from pre-existing models, transfer learning can enhance the accuracy of models trained on limited HSI data. On the other hand, active learning involves selecting and annotating informative samples to improve the effectiveness of models trained on small HSI datasets. By combining the two techniques, transfer learning and active learning can address the bottleneck of limited HSI data and enable the development of robust 3D-CNN models capable of accurately classifying HSI data.

The computational complexity of 3D-CNN models is a barrier for their deployment in the field, for example, in edge computing devices or even as part of an embedded system in UAVs or agricultural equipment. Real-time diagnosis could significantly enable farmers to quickly detect and respond to disease outbreaks, which can help to prevent the spread of disease and reduce crop losses. This requires, however, researchers to explore ways to optimize their models for speed and efficiency, particularly their memory-footprint. This can be achieved through advancing the capabilities of single board computers, particularly their GPU and AI accelerators, on the one side, as well as developing lightweight models on the other side, for example, by reduction of used bands or grouping of bands into indices. Identifying the spectral bands that are most informative for detecting a varied range of diseases and defects would play an important role in reducing the model's training computational time, decreasing the number of parameters, achieving higher accuracy and generating a more lightweight model.

Despite the valuable insights hyperspectral imaging provides regarding the condition and health of crops, deploying this technology is relatively costly compared to RGB and multispectral imaging technologies. The higher expense is primarily attributed to the increased processing power required to analyze the HSI data and the extensive range of spectrum offered by hyperspectral cameras.

One approach to consider for cost reduction is the limitation of the number of bands in hyperspectral cameras, as not all spectral bands may be equally crucial for disease and defect detection. Therefore, instead of having an imaging device that supports full range of wavelengths from the visible to the infrared spectrum, we can deploy imaging systems that work with essential wavelengths rather than hundreds of spectral bands. In this respect, some companies already provide the facility to design customized multispectral systems, which work with bands that had been identified to be the most discriminative (see Hamila et al. [22] for a 3D-CNN model training advantage of this approach). Therefore, developing methods for identifying the most informative spectral bands would also be cost beneficial. Furthermore, by leveraging Machine Learning as a Service (MLaaS) platforms [48], the need for required infrastructure to process HSI data using 3D-CNN can be eliminated. Additionally, leading MLaaS providers can integrate pre-built 3D-CNN models into their offerings for training HSI data, thereby reducing the time and effort required for model development, especially for individuals in the agriculture domain who may have limited machine learning expertise.

In summary, the application of 3D-CNNs with hyperspectral data for disease and defective crop detection is a promising research area with several open research questions. Collecting and sharing HSI data on scale, identifying informative spectral bands, developing transfer and active learning techniques, and implementing light-weight architectures are some of the key research areas that can be considered for future work. The advancement of this research will lead to more accurate and efficient disease and defective crop detection, which can have significant impacts on the agricultural industry.

CRediT authorship contribution statement

Nooshin Noshiri: Conceptualization, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. **Michael A. Beck:** Writing – original draft, Writing – review & editing. **Christopher P. Bidinosti:** Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – review & editing. **Christopher J. Henry:** Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgement

References

[1] L. Alzubaidi, J. Bai, A. Al-Sabaawi, J. Santamaría, A.S. Albahri, B.S.N. Al-dabbagh, M.A. Fadhel, M. Manoufali, J. Zhang, A.H. Al-Timemy, Y. Duan, A. Abdullah, L. Farhan, Y. Lu, A. Gupta, F. Albu, A. Abbosh, Y. Gu, A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications, J. Big Data 10 (2023) 46, https://doi.org/10.1186/s40537-023-00727-2.

[2] J. Avbelj, Spectral information retrieval for sub-pixel building edge detection, ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci. I-7 (2012) 61–66, https://doi.org/10.5194/isprsannals-I-7-61-2012.

[3] M.A. Beck, C.P. Bidinosti, C.J. Henry, M. Ajmani, The terrabyte client: providing access to terabytes of plant data, arXiv:2203.13691 [abs], https://doi.org/10.48550/arXiv.2203.13691, 2022.

[4] M.A. Beck, C.Y. Liu, C.P. Bidinosti, C.J. Henry, C.M. Godee, M. Ajmani, An embedded system for the automated generation of labeled plant images to enable machine learning applications in agriculture, PLoS ONE 15 (2020) e0243923, https://doi.org/10.1371/journal.pone.0243923.

[5] J.A. Benediktsson, P. Ghamisi, Spectral-Spatial Classification of Hyperspectral Remote Sensing Images, Artech House, 2015.

[6] A. Bojeri, F. Melgani, G. Giannotta, G. Ristorto, G. Guglieri, J.M. Junior, Automatic crop rows segmentation for multispectral aerial imagery, in: Mediterranean and Middle-East Geoscience and Remote Sensing Symposium (M2GARSS), IEEE, 2022, pp. 78–81.

[7] T. Boston, A. Van Dijk, P.R. Larraondo, R. Thackway, Comparing cnns and random forests for landsat image segmentation trained on a large proxy land cover dataset, Remote Sens. 14 (2022) 3396, https://doi.org/10.3390/rs14143396.

[8] P. Bunting, D. Clewley, R.M. Lucas, S. Gillingham, The Remote Sensing and GIS Software Library (RSGISLib), Computers and Geosciences, vol. 62, 2014, pp. 216–226.

[9] X. Cao, B. Ji, Y. Ji, L. Wang, L. Jiao, Hyperspectral image classification based on filtering: a comparative study, J. Appl. Remote Sens. 11 (2017) 035007, https://doi.org/10.1117/1.JRS.11.035007.

[10] Y. Cao, P. Yuan, H. Xu, J.F. Martínez-Ortega, J. Feng, Z. Zhai, Detecting asymptomatic infections of rice bacterial leaf blight using hyperspectral imaging and 3-dimensional convolutional neural network with spectral dilated convolution, Front. Plant Sci. 13 (2022), https://doi.org/10.3389/fpls.2022.963170.

[11] Centre National D'Etudes Spatiales (CNES), Orfeo toolbox, https://www.orfeo-toolbox.org/download/, 2023.

[12] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, Smote: synthetic minority over-sampling technique, J. Artif. Intell. Res. 16 (2002) 321–357, https://doi.org/10.1613/jair.953.

[13] S.Y. Chen, M.F. Chiu, X.W. Zou, Real-time defect inspection of green coffee beans using nir snapshot hyperspectral imaging, Comput. Electron. Agric. 197 (2022) 106970, https://doi.org/10.1016/j.compag.2022.106970.

[14] M. Dalla Mura, J.A. Benediktsson, B. Waske, L. Bruzzone, Morphological attribute profiles for the analysis of very high resolution images, IEEE Trans. Geosci. Remote Sens. 48 (2010) 3747–3762, https://doi.org/10.1109/TGRS.2010.2048116.

[15] B.B. Damodaran, N. Courty, S. Lefèvre, Sparse hilbert schmidt independence criterion and surrogate-kernel-based feature selection for hyperspectral image classification, IEEE Trans. Geosci. Remote Sens. 55 (2017) 2385–2398, https://doi.org/10.1109/TGRS.2016.2642479.

[16] A. Datta, S. Ghosh, A. Ghosh, Combination of clustering and ranking techniques for unsupervised band selection of hyperspectral images, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 8 (2015) 2814–2823, https://doi.org/10.1109/JSTARS.2015.2428276.

[17] S. Duggal, Surveying, vol. 2, Tata McGraw-Hill Education, 2013.

[18] H. Fırat, M.E. Asker, D. Hanbay, Classification of hyperspectral remote sensing images using different dimension reduction methods with 3d/2d cnn, Remote Sens. Appl. Soc. Environ. 25 (2022) 100694, https://doi.org/10.1016/j.rsase.2022.100694.

[19] T. Gao, A.K.N. Chandran, P. Paul, H. Walia, H. Yu, Hyperseed: an end-to-end method to process hyperspectral images of seeds, Sensors 21 (2021) 8184, https://doi.org/10.3390/s21248184.

[20] P. Ghamisi, M.S. Couceiro, J.A. Benediktsson, A novel feature selection approach based on fodpso and svm, IEEE Trans. Geosci. Remote Sens. 53 (2014) 2935–2947, https://doi.org/10.1109/TGRS.2014.2367010.

[21] T. Gill, S.K. Gill, D.K. Saini, Y. Chopra, J.P. de Koff, K.S. Sandhu, A comprehensive review of high throughput phenotyping and machine learning for plant stress phenotyping, Phenomics 2 (2022) 156–183, https://doi.org/10.1007/s43657-022-00048-z.

[22] O. Hamila, C.J. Henry, O.I. Molina, C.P. Bidinosti, M.A. Henriquez, Fusarium head blight detection, spikelet estimation, and severity assessment in wheat using 3d convolutional neural networks, arXiv:2303.05634 [abs], https://doi.org/10.3390/rs11222658, 2023.

[23] H. He, Y. Bai, E.A. Garcia, S. Li, Adasyn: adaptive synthetic sampling approach for imbalanced learning, in: Proceedings of International Joint Conference on Neural Networks (IJCNN), IEEE, Hong Kong, 2008, pp. 1322–1328.

[24] V. Henrich, E. Götze, A. Jung, C. Sandow, D. Thürkow, C. Gläßer, Development of an online indices database: motivation, concept and implementation, in: 6th EARSeL Imaging Spectroscopy Sig Workshop Innovative Tool for Scientific and Commercial Environment Applications, Tel Aviv, Israel, 2009, pp. 16–18, https://www.indexdatabase.de/info/credits.php.

[25] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceedings of Computer Vision and Pattern Recognition Conference (CVPR), IEEE, Salt Lake City, UT, USA, 2018, pp. 7132–7141.

[26] A. Ifarraguerri, M.W. Prairie, Visual method for spectral band selection, IEEE Geosci. Remote Sens. Lett. 1 (2004) 101–106, https://doi.org/10.1109/LGRS.2003.822879.

[27] T. Imbiriba, J.C.M. Bermudez, C. Richard, J.Y. Tourneret, Band selection in rkhs for fast nonlinear unmixing of hyperspectral images, in: Proceedings of European Signal Processing Conference (EUSIPCO), IEEE, Nice, France, 2015, pp. 1651–1655.

[28] Y. Jia, Y. Shi, J. Luo, H. Sun, Y–net: identification of typical diseases of corn leaves using a 3d–2d hybrid cnn model combined with a hyperspectral image band selection module, Sensors 23 (2023) 1494, https://doi.org/10.3390/s23031494.

[29] Y. Jiang, C. Li, Convolutional neural networks for image-based high-throughput plant phenotyping: a review, Plant Phenom. (2020), https://doi.org/10.34133/2020/4152816.

[30] X. Jin, L. Jie, S. Wang, H.J. Qi, S.W. Li, Classifying wheat hyperspectral pixels of healthy heads and fusarium head blight disease using a deep neural network in the wild field, Remote Sens. 10 (2018) 395, https://doi.org/10.3390/rs10030395.

[31] D.S. Joseph, P.M. Pawar, R. Pramanik, Intelligent plant disease diagnosis using convolutional neural network: a review, Multimed. Tools Appl. 82 (2022) 1–67, https://doi.org/10.1007/s11042-022-14004-6.

[32] V.H. Josephine, A. Nirmala, V.L. Alluri, Impact of Hidden Dense Layers in Convolutional Neural Network to Enhance Performance of Classification Model, IOP Conf. Ser., Mater. Sci. Eng. (2021) 012007, https://doi.org/10.1088/1757-899X/1131/1/012007, IOP Publishing.

[33] D.H. Jung, J.D. Kim, H.Y. Kim, T.S. Lee, H.S. Kim, S. Park, A hyperspectral data 3d convolutional neural network classification model for diagnosis of gray mold disease in strawberry leaves, Front. Plant Sci. 13 (2022) 620, https://doi.org/10.3389/fpls.2022.837020.

[34] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T.Y. Liu, Lightgbm: a highly efficient gradient boosting decision tree, Adv. Neural Inf. Process. Syst. 30 (2017).

[35] J. Khodr, R. Younes, Dimensionality reduction on hyperspectral images: a comparative review based on artificial datas, in: Proceedings of International Congress on Image and Signal Processing, IEEE, Shanghai, China, 2011, pp. 1875–1883.

[36] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Commun. ACM 60 (2017) 84–90, https://doi.org/10.1145/3065386.

[37] P. Kumar, D.K. Gupta, V.N. Mishra, R. Prasad, Comparison of support vector machine, artificial neural network, and spectral angle mapper algorithms for crop classification using liss iv data, Int. J. Remote Sens. 36 (2015) 1604–1617, https://doi.org/10.1080/2150704X.2015.1019015.

[38] L3Harris Geospatial, Envi., https://www.l3harrisgeospatial.com/Software-Technology/ENVI, 2023.

[39] M. Lennon, G. Mercier, L. Hubert-Moy, Nonlinear filtering of hyperspectral images with anisotropic diffusion, in: Proceedings of International Geoscience and Remote Sensing Symposium, IEEE, Toronto, ON, Canada, 2002, pp. 2477–2479.

[40] J. Li, H. Wang, A. Zhang, Y. Liu, Semantic segmentation of hyperspectral remote sensing images based on pse-unet model, Sensors 22 (2022) 9678, https://doi.org/10.3390/s22249678.

[41] J.m. Li, Y.t. Qian, Clustering-based hyperspectral band selection using sparse non-negative matrix factorization, J. Zhejiang Univ. Sci. C 12 (2011) 542–549, https://doi.org/10.1631/jzus.C1000304.

[42] Z. Liu, J. Jiang, X. Qiao, X. Qi, Y. Pan, X. Pan, Using convolution neural network and hyperspectral image to identify moldy peanut kernels, Lebensm.-Wiss. Technol. 132 (2020) 109815, https://doi.org/10.1016/j.lwt.2020.109815.

[43] G. Luo, G. Chen, L. Tian, K. Qin, S.E. Qian, Minimum noise fraction versus principal component analysis as a preprocessing step for hyperspectral imagery denoising, Can. J. Remote Sens. 42 (2016) 106–116, https://doi.org/10.1080/07038992.2016.1160772.

[44] B. Mojaradi, H. Emami, M. Varshosaz, S. Jamali, A novelband selection method for hyperspectral data analysis, Proc. of Int. Arch. Photogramm., Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. 447 (2008) 451.

[45] K. Nagasubramanian, S. Jones, A.K. Singh, A. Singh, B. Ganapathysubramanian, S. Sarkar, Explaining hyperspectral imaging based plant disease identification: 3d cnn and saliency maps, arXiv:1804.08831 [abs], 2018.

[46] NASA Ocean Biology Processing Group (OBPG), Seadas, https://seadas.gsfc.nasa.gov/downloads/, 2022.

[47] C. Nguyen, V. Sagan, M. Maimaitiyiming, M. Maimaitijiang, S. Bhadra, M.T. Kwasniewski, Early detection of plant viral disease using hyperspectral imaging and deep learning, Sensors 21 (2021) 742, https://doi.org/10.3390/s21030742.

[48] N. Noshiri, M. Khorramfar, T. Halabi, Machine learning-as-a-service performance evaluation on multi-class datasets, in: International Conference on Smart Internet of Things (SmartIoT), IEEE, 2021, pp. 332–336.

[49] A. Plaza, G. Martín, J. Plaza, M. Zortea, S. Sánchez, Recent Developments in Endmember Extraction and Spectral Unmixing, Springer, Berlin Heidelberg, 2011, pp. 235–267.

[50] R. Pourdarbani, S. Sabzi, M. Dehghankar, M.H. Rohban, J.I. Arribas, Examination of lemon bruising using different cnn-based classifiers and local spectral-spatial hyperspectral imaging, Algorithms 16 (2023) 113, https://doi.org/10.3390/a16020113.

[51] C. Qi, M. Sandroni, J.C. Westergaard, E.H.R. Sundmark, M. Bagge, E. Alexandersson, J. Gao, In-field classification of the asymptomatic biotrophic phase of potato late blight based on deep learning and proximal hyperspectral imaging, Comput. Electron. Agric. 205 (2023) 107585, https://doi.org/10.1016/j.compag.2022.107585.

[52] S. Qiao, Q. Wang, J. Zhang, Z. Pei, Detection and classification of early decay on blueberry based on improved deep residual 3d convolutional neural network in hyperspectral images, Sci. Program. 2020 (2020), https://doi.org/10.1155/2020/8895875.

[53] J. Qin, T.F. Burks, M.A. Ritenour, W.G. Bonn, Detection of citrus canker using hyperspectral reflectance imaging with spectral information divergence, J. Food Eng. 93 (2009) 183–191, https://doi.org/10.1016/j.jfoodeng.2009.01.014.

[54] W.G. Rees, Physical Principles of Remote Sensing, 3 ed., Cambridge University Press, 2013.

[55] Resonon Inc, Spectronon software, https://resonon.com/software, 2023.

[56] D.M. Rizzo, M. Lichtveld, J.A.K. Mazet, E. Togami, S.A. Miller, Plant health and its effects on food safety and security in a one health framework: four case studies, One Health Outlook 3 (2021), https://doi.org/10.1186/s42522-021-00038-7.

[57] L.C.B.D. Santos, S.J.F. Guimaraes, J.A.D. Santos, Efficient unsupervised band selection through spectral rhythms, IEEE J. Sel. Top. Signal Process. 9 (2015) 1016–1025, https://doi.org/10.1109/JSTSP.2015.2405902.

[58] S.S. Sawant, M. Prabukumar, A survey of band selection techniques for hyperspectral image classification, J. Spectr. Imag. 9 (2020), https://doi.org/10.1255/jsi.2020.a5.

[59] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-cam: visual explanations from deep networks via gradient-based localization, Int. J. Comput. Vis. 128 (2020) 336–359, https://doi.org/10.1007/s11263-019-01228-7.

[60] M.S. Shaikh, K. Jaferzadeh, B. Thörnberg, J. Casselgren, Calibration of a hyperspectral imaging system using a low-cost reference, Sensors 21 (2021) 3738, https://doi.org/10.3390/s21113738.

[61] V. Sharma, A. Diba, T. Tuytelaars, L. Van Gool, Hyperspectral CNN for image classification & band selection, with application to face recognition, Technical Report, KU Leuven, 2016.

[62] A. Shi, A. Gao, Z. He, M. Li, L. Xu, A hyperspectral band selection based on game theory and differential evolution algorithm, Int. J. Smart Sens. Intell. Syst. 9 (2016) 1971–1990, https://doi.org/10.21307/ijssis-2017-948.

[63] A. Signoroni, M. Savardi, A. Baronio, S. Benini, Deep learning meets hyperspectral image analysis: a multidisciplinary review, J. Imaging 5 (2019) 52, https://doi.org/10.3390/jimaging5050052.

[64] K. Simonyan, A. Vedaldi, A. Zisserman, Deep inside convolutional networks: visualising image classification models and saliency maps, arXiv:1312.6034 [abs], https://doi.org/10.48550/arXiv.1312.6034, 2013.

[65] H. Su, Y. Sheng, P. Du, K. Liu, Adaptive affinity propagation with spectral angle mapper for semi-supervised hyperspectral band selection, Appl. Opt. 51 (2012) 2656–2663, https://doi.org/10.1364/AO.51.002656.

[66] H. Su, H. Yang, Q. Du, Y. Sheng, Semisupervised band clustering for dimensionality reduction of hyperspectral imagery, IEEE Geosci. Remote Sens. Lett. 8 (2011) 1135–1139, https://doi.org/10.1109/LGRS.2011.2158185.

[67] D.W. Sun, Hyperspectral Imaging for Food Quality Analysis and Control, Elsevier, 2010.

[68] K. Sun, X. Geng, L. Ji, An efficient unsupervised band selection method based on an autocorrelation matrix for a hyperspectral image, Int. J. Remote Sens. 35 (2014) 7458–7476, https://doi.org/10.1080/01431161.2014.968686.

[69] K. Sun, X. Geng, L. Ji, A new sparsity-based band selection method for target detection of hyperspectral image, IEEE Geosci. Remote Sens. Lett. 12 (2014) 329–333, https://doi.org/10.1109/LGRS.2014.2337957.

[70] L. Sun, X. Song, H. Guo, G. Zhao, J. Wang, Patch-wise semantic segmentation for hyperspectral images via a cubic capsule network with emap features, Remote Sens. 13 (2021) 3497, https://doi.org/10.3390/rs13173497.

[71] W. Sun, Q. Du, Hyperspectral band selection: a review, IEEE Geosci. Remote Sens. Mag. 7 (2019) 118–139, https://doi.org/10.1109/MGRS.2019.2911100.

[72] W. Sun, L. Tian, Y. Xu, D. Zhang, Q. Du, Fast and robust self-representation method for hyperspectral band selection, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 10 (2017) 5087–5098, https://doi.org/10.1109/JSTARS.2017.2737400.

[73] H.H. Tan, K.H. Lim, Vanishing gradient mitigation with deep learning neural network optimization, in: Proceedings of International Conference on Smart Computing & Communications (ICSCC), IEEE, Sarawak, Malaysia, 2019, pp. 1–4.

[74] Y. Tarabalka, J.A. Benediktsson, J. Chanussot, J.C. Tilton, Multiple spectral–spatial classification approach for hyperspectral data, IEEE Trans. Geosci. Remote Sens. 48 (2010) 4122–4132, https://doi.org/10.1109/TGRS.2010.2062526.

[75] J. Tschannerl, J. Ren, J. Zabalza, S. Marshall, Segmented autoencoders for unsupervised embedded hyperspectral band selection, in: Proceedings of European Workshop on Visual Information Processing (EUVIP), IEEE, Tampere, Finland, 2018, pp. 1–6.

[76] M. Uzair, N. Jamil, Effects of hidden layers on the efficiency of neural networks, in: IEEE 23rd International Multitopic Conference (INMIC), IEEE, 2020, pp. 1–6.

[77] C. Vaiphasa, Consideration of smoothing techniques for hyperspectral remote sensing, ISPRS J. Photogramm. Remote Sens. 60 (2006) 91–99, https://doi.org/10.1016/j.isprsjprs.2005.11.002.

[78] M. Vidal, J.M. Amigo, Pre-processing of hyperspectral images. essential steps before image analysis, Chemom. Intell. Lab. Syst. 117 (2012) 138–148, https://doi.org/10.1016/j.chemolab.2012.05.009.

[79] D. Wan, R. Lu, S. Wang, S. Shen, T. Xu, X. Lang, Yolo-hr: improved yolov5 for object detection in high-resolution optical remote sensing images, Remote Sens. 15 (2023) 614, https://doi.org/10.3390/rs15030614.

[80] C. Wang, B. Liu, L. Liu, Y. Zhu, J. Hou, P. Liu, X. Li, A review of deep learning used in the hyperspectral image analysis for agriculture, Artif. Intell. Rev. 54 (2021) 5205–5253, https://doi.org/10.1007/s10462-021-10018-y.

[81] L. Wang, X. Jia, Y. Zhang, A novel geometry-based feature-selection technique for hyperspectral imagery, IEEE Geosci. Remote Sens. Lett. 4 (2007) 171–175, https://doi.org/10.1109/LGRS.2006.887142.

[82] C. Yang, Y. Tan, L. Bruzzone, L. Lu, R. Guan, Discriminative feature metric learning in the affinity propagation model for band selection in hyperspectral images, Remote Sens. 9 (2017) 782, https://doi.org/10.3390/rs9080782.

[83] J. Yang, Z. Jin, J. Yang, Non-linear techniques for dimension reduction, in: Encyclopedia of Biometrics, Springer, 2009, pp. 1003–1007.

[84] J. Yin, Y. Wang, Z. Zhao, Optimal band selection for hyperspectral image classification based on inter-class separability, in: Proceedings of Symposium on Photonics and Optoelectronics, IEEE, Chengdu, China, 2010, pp. 1–4.

[85] S. Yin, Y. Wang, Y.H. Yang, A novel image-dehazing network with a parallel attention block, Pattern Recognit. 102 (2020) 107255, https://doi.org/10.1016/j.patcog.2020.107255.

[86] H. Zhai, H. Zhang, L. Zhang, P. Li, Squaring weighted low-rank subspace clustering for hyperspectral image band selection, in: Proceedings of International Geoscience and Remote Sensing Symposium (IGARSS), IEEE, Beijing, China, 2016, pp. 2434–2437.

[87] Y. Zhan, D. Hu, H. Xing, X. Yu, Hyperspectral band selection based on deep convolutional neural network and distance density, IEEE Geosci. Remote Sens. Lett. 14 (2017) 2365–2369, https://doi.org/10.1109/LGRS.2017.2765339.

[88] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, A. Torralba, Learning deep features for discriminative localization, in: Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Las Vegas, NV, USA, 2016, pp. 2921–2929.