Smart Agricultural Technology 12 (2025) 101183

Contents lists available at ScienceDirect

Smart Agricultural Technology

journal homepage: www.journals.elsevier.com/smart-agricultural-technology

Multiscale transformer-based network for rangeland plant classification used in pasture scoring

Zakieh Alizadehsani [a], Oliver Hensel [a,*], Abozar Nasirahmadi [a,b]

[a] Department of Agricultural and Biosystems Engineering, University of Kassel, Witzenhausen D-37213, Germany
[b] Department of Energy and Technology, Swedish University of Agricultural Sciences, Box 7032, Uppsala 75007, Sweden

ARTICLE INFO

Keywords:
Machine learning
Vision transformers
Rangeland ecosystem
Multiscale learning

ABSTRACT

Rangeland ecosystems have been sources for pastoral communities. However, traditional seasonal mobility patterns are disrupted by climate change, requiring more dynamic, data-driven plant-based rangeland assessment. In this study, we propose amultiscale transformer-based network to address the challenge of automatically classifying rangeland plant species for livestock pasture scoring in Africa, given the complex environments and limited data. Accurately distinguishing similar plants with varying livestock utility is important for sustainable management. This study investigated Vision Transformers, known for multiscale features important for fine-grained visual differentiating. The initial comparative analysis of ViT, DEiT, and Swin Transformer models demonstrated the promise of Swin architecture. Building on this, we introduce a Multiscale Swin Transformer model incorporating multiscale feature fusion and weighting mechanism to enhance plant image classification. The model combines global and fine feature extraction, followed by fusion module. Early features capture local patterns (e.g., leaf), and later layers capture semantic information (e.g., general morphology). The proposed Multiscale approach utilizing a weighted decoder provides better performance improvements over the Swin base model, achieving 89.71 % accuracy compared to 88.0%, demonstrating that fusing features at different scales leads to better recognition. Moreover, analysis of the collected data shows class imbalance, including dominance of invasive species and useful herbs, sparse representation of rare unuseful (e.g., poisonous) and other sparse useful livestock forage species. This highlights an essential need for systematic data collection and optimization strategies, like synthetic image generation, to mitigate limitations and improve model generalization.

1. Introduction

Plant-based livestock feeding, using natural environments like rangeland ecosystems, is an important source for livestock management in Africa [14]. Recently, climate change has been changing the seasonal pattern of the composition of plants [36]. Since rangeland-based livestock pasture selection is highly dependent on using natural environments, information and communications technology (ICT) tools can help to collect data for more informative understanding and automatic analyzing of the environment [19]. In this context, ICT tools can provide an opportunity for computer vision to provide automated bulk image analysis from plant species images received from local communities.

Plant classification models often struggle with challenges in natural environments (e.g., inter-class image variations)[49]. For African species, this is relevant considering insufficient local data and scale variation. These issues arise from difference between training data and real-world images, especially concerning variations in devices and distances [17]. Moreover, environmental factors all contribute to the difficulty of classifying plants in uncontrolled environments such as rangeland with different natural resources. Additionally, applying state-of-the-art computer vision and transfer learning models trained on public datasets with limited African data [10], and therefore lacking appropriate representation of African plant species presents notable challenges. In this work, our investigation focused on two key areas: data characterization and machine learning technique evaluation. Considering the collected data, the analysis showed a natural species distribution imbalance, with herbs and invasive species [28] dominating the majority classes. In contrast, minority classes, including rare or geographically isolated forage species, are underrepresented. These minority classes, despite their sparse distribution or challenging accessibility (e.g., mountainous regions), are important for livestock health.

Based on identified data characteristics and investigation of transfer-

* Corresponding author.
E-mail address: agrartechnik@uni-kassel.de (O. Hensel).

https://doi.org/10.1016/j.atech.2025.101183
Received 5 May 2025; Received in revised form 9 July 2025; Accepted 9 July 2025
Available online 10 July 2025

based models performance for rangeland plant species classification. This study first investigated a comparative analysis of well-known transformers for plant classification and afterward introduces a Multi-scale Swin Transformer-based learning model to address these challenges at various stages of the proposed learning architecture. As the main contribution, this study takes advantage of transformer-based architecture and multi-scale feature fusion. Strong feature extraction through feature fusion methods is utilized to integrate information from varying levels of plant image detail, enabling accurate species identification despite inter-class diversity resulting from seasonal changes and different growth stages. Image fusion methods can be broadly categorized into mathematical transformations in spatial or transform domains and deep learning-based methods [52], employing neural networks to learn and combine features. In this study, deep learning-based methods were used to fuse hierarchical layers and combine local botanical features with global plant morphology while handling the natural imbalance in plant species distribution. Within this framework, to address the challenge of class imbalance in our plant-based classification task, we implemented a weighted decoder architecture. Considering the imbalanced distribution of our African plant dataset, the proposed multiscale approach with weighted decoder achieved better performance than the Swin base model in terms of classification metrics. Moreover, analysis of model performances indicates the Multiscale Swin-*B*+WD demonstrates a balance between classification metrics (83.45 % at >0.5 threshold) and data preservation (73.0 %).

The rest of this paper is organized into the following sections. Section 2 gives an overview of related works. Section 3 proposed the methodology. The experimental results are represented in Section 4. Finally, Section 5 and 6 is related to the discussion and conclusion, and future work.

2. Related work

The development of machine learning models and their increasing accuracy and efficiency have led to an increasing interest in computer vision-based identification of plant species since last decades [31]. Existing works can be classified into three main categories, including non-learning methods, such as domain knowledge-based and morphometric methods [23], Machine learning, and Deep learning methods, which will be described in the rest of this section.

- **Image processing**: Several automated methods have been proposed, aiming to turn plant image analysis problems into computer vision tasks. In traditional methods using image processing, researchers take inputs like plant features (e.g., color, texture) and define a set of architectures that describe the relationship between plant features using mathematical models. For example, color-based approaches [37] have been utilized to isolate specific plant parts, enabling targeted analysis. However, considerable spectral overlaps exist across plant species and within developmental stages, making accurate differentiation challenging. Consequently, complementing color-based approaches, texture analysis, and image filtering [46] enable the discrimination of plant features through the analysis of surface textures, effectively addressing the challenges of spectral overlap. For example, two green leaves might have similar color values, but their surface textures (e.g., smooth) can be used to differentiate them. However, texture analyses are affected by inconsistencies in illumination and the presence of noise, and image filtering methods like Gabor filters require precise parameter adjustments (e.g., scale, phase), making them highly sensitive to tuning [27]. While shape quantification algorithms [35], Fourier descriptors [50], and edge detection [8] offer alternative approaches, the highly irregular edges common to plant leaves introduce noise sensitivity during contour tracing, directly impacting the accuracy of Fourier descriptors.
- **Machine learning**: Increasing accuracy has become important when it comes to plant analysis (i.e., monitoring plants). Machine learning models are able to identify complex, nonlinear patterns in data, and they can also adapt to new data and improve independently, so it has become a better option. The proposed machine learning models generally utilize linear algebra and probability theory to generate predictions from labeled (supervised) [1] or unlabeled (unsupervised) [51] datasets. Support Vector Machines (SVMs) and Random Forests have been widely employed [1,29], leveraging handcrafted features like leaf shape, texture, and color of the target plant dataset. For example, Larese et al. [32] have used SVM to classify plant species based on leaf vein patterns. Random Forests have demonstrated effectiveness in handling high-dimensional feature spaces derived from leaf morphology [32]. Nevertheless, both approaches rely heavily on accurate image feature extraction and preprocessing, including segmentation and feature extraction. Errors in these steps can propagate through the classification pipeline, consequently reducing accuracy. Moreover, while Random Forests, as ensemble learning method, handle high-dimensional data, the complexity of leaf morphology can lead to overfitting, mostly with limited training data.
- **Deep learning**: Deep learning models, in contrast to traditional machine learning, automate the process of feature extraction using neural networks. Additionally, deep learning-based models extract more reliable features when faced with large and complex datasets. Among deep learning-based models, Convolutional Neural Networks (CNNs), have advanced plant classification accuracy [33,39]. The methods based on CNN's learn hierarchical features directly from raw image data and use CNN for feature extraction, eliminating the need for manual feature engineering. This capability allows them to capture important variations in plant morphology and texture, leading to better performance compared to traditional machine learning methods. For instance, pre-trained CNN models, such as ResNet [22], AlexNet [30], and Inception [41], have been utilized for plant species identification, achieving state-of-the-art results on large plant image datasets. These models can effectively handle challenges like variations in plant species and image quality, making them strong for real-world applications. Some models use stated pre-trained architectures, including [45] (AlexNet), [15] (VGGNet), and [6,39] (ResNet). There are also studies that compare their methodologies to these established models. As an example, Wei et al. [48] proposed D-Leaf, a CNN-based method for identifying plant species automatically, and evaluated it against pre-trained and fine-tuned AlexNet. Ghazi et al. [16] employed pre-trained AlexNet, GoogLeNet [41], and VGGNet [38] for plant species identification to compare the performance of pre-trained AlexNet, GoogLeNet, and VGGNet. The emergence of transformers, which originated from natural language processing, has reshaped computer vision models; this paradigm is followed in plant-based vision tasks.

Recently, there has been interest in the use of abstracted features like attention-mechanism [4] and transformers for plant-related tasks in recent years [26], such as ViT [11] and Swin [34], has improved their ability to capture global context, handle high-resolution images, and provide superior feature representation, making them preferred for automation-based tasks. For instance, Gole et al. [18] utilized ViT for early plant disease detection, substituting the MLP with an Inception module in a novel block. However, ViTs, in their basic form, divide images into fixed-size patches. This makes ViT-based less resilient to scale variations, which are common in natural plant images (e.g., leaves at different distances, plants of varying sizes). Swin Transformers, with their hierarchical structure and window-based attention, are better at capturing multiscale features. To this end, Swin was employed in several works [11,25,53]) by Guo et al. [20] to determine the type and extent of plant diseases. Moreover, some deep learning methods primarily extract abstract features using deeper layers or attention mechanisms. However,

obtaining detailed information across various image scales, as demonstrated in some change detection studies, can offer different levels of granularity [54]. While plant species are typically studied in controlled environments, this study focuses on less accessible African species in a real-world environment. Within this framework, based on the data characteristics, we propose a Multiscale architecture allowing extracts both global and fine-grained plant features for better identification.

3. Methodology

This study started with an evaluation of three established transformer-based architectures: ViT (Vision Transformer) [11], DeiT [42] (Data-Efficient Image Transformer), and Swin [34]. These models were assessed on the plant classification task to provide a baseline and understand the strengths and weaknesses of each architecture. As discussed in Section 4.2, the Swin Transformer demonstrated promising performance, mostly based on its hierarchical design, which captures multiscale features. This observation motivated its selection as the foundation for our proposed model. Subsequently, the following sections explain how the dataset was collected, and then the proposed Multiscale model for rangeland-based plant identification will be described.

3.1. Data acquisition and optimization

Deep learning-based plant classification, like other image classification tasks, typically requires vast amounts of labeled data for effective training. To this end, the first step is the collection of a well-annotated dataset [2]. The rise of internet access and smartphone technology has opened up possibilities for using mobile-based surveys in rangeland environments [43]. This approach allows for monitoring and gathering data (e.g., images) in natural environments that are hard to reach in different seasons. Furthermore, mobile sensors can capture metadata, including location information via GPS [2]. Building on this technological foundation, this study collected data in Kenya's arid and semi-arid Marsabit County through a systematic framework designed to capture a diverse and representative dataset of rangeland plants. The data collection process began with training sessions for local photographers to ensure best practices for botanical image capture, followed by selection of target grazing areas through a collaborative approach involving local herders who identified and named plants based on their ecological and economic importance. This participatory method ensured that the dataset reflects the ecologically and economically important vegetation as recognized by those with local knowledge of rangeland management. In addition the identified species were subsequently validated using resources such as the [16] and "African Plants a Photo Guide" [3,21] to ensure scientific accuracy.

The data collection process resulted in eight rangeland plant classes demonstrated in Fig. 1, encompassing four useful forage species that serve as important food sources for livestock (*Acacia, Bissar, Acacia reficiens, Maerua edulis*, and Yabah), one less useful species (*Boscia coriacea*), and problematic species including the invasive *Prosopis juliflora* and the poisonous *Desert rose* that pose significant threats to livestock health and productivity [47]. The selected species are referred to by their scientific and local names.To ensure temporal representativeness, the selected areas and plant species were systematically sampled and revisited during both dry and wet seasons.

Considering image quality directly impacts model performance, technical standards were implemented throughout the data collection process. The protocol required field photographers to focus on distinct, individual plants in each image, and capture the entire plant structure. Moreover, in the quality filter step, repetitive images, excluding non-plant objects such as buildings, trash, or other man-made structures have been removed to avoid unwanted artifacts in the training data. Images were captured exclusively during daylight hours under proper lighting conditions. Furthermore, photographs were taken from varying

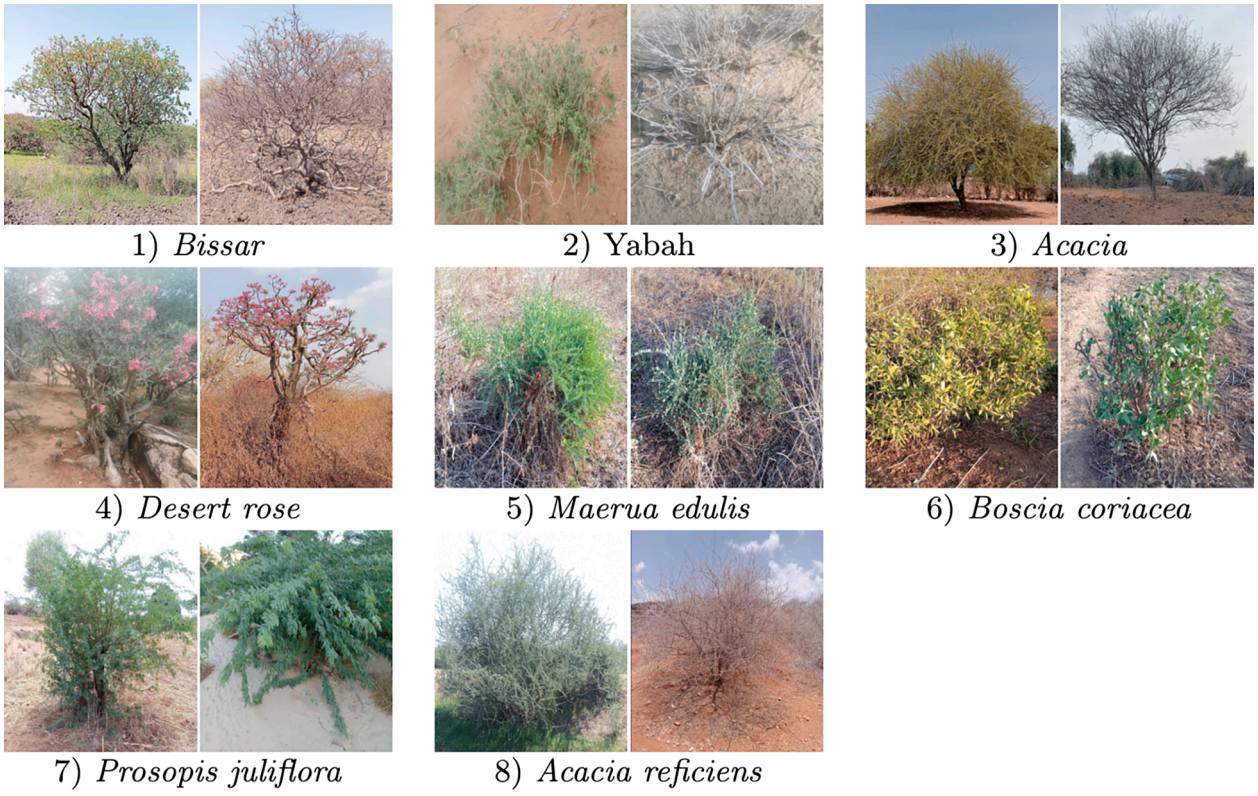

Fig. 1. Plant species collected to represent the diverse growth forms and seasonal availability of livestock fodder (e.g., shrubs, herbs) in arid and semi-arid regions. These plant species include both useful and unuseful (e.g., invasive) varieties, representing the plant distribution in the rangeland environment.

distances and perspectives to align with the visual identification needs of local communities, ensuring that the resulting dataset would support practical applications in real-world rangeland management scenarios.

In this study, a total of 4275 image samples were split into a 60–20–20 ratio to ensure sufficient data for testing. Considering the imbalanced distribution of the data, systematic data augmentation has been applied and the weighted decoder has been implemented based on the characteristics of the imbalanced datasets (section 3.2). The train dataset includes 2565 images with a naturally imbalanced distribution demonstrating real-world rangeland conditions, Yabah as the majority class with 549 samples (21.40 %), followed by *Prosopis juliflora* with 435 samples (16.96 %), *Boscia coriacea* with 399 samples (15.56 %), *Maerua edulis* with 331 samples (12.90 %), *Acacia* with 313 samples (12.20 %), *Bissar* with 293 samples (11.42 %), *Acacia reficiens*with 168 samples (6.55 %), and *Desert rose* as the least represented with only 77 samples (3.00 %), resulting in a plant imbalance ratio of 1:0.14. Following systematic data augmentation, the augmented dataset of 3973 samples achieved improved balance, with plant distribution including Yabah with 699 samples (17.59 %), *Bissar* enhanced to 607 samples (15.28 %), *Prosopis juliflora* with 538 samples (13.54 %), *Boscia coriacea* reaching 512 samples (12.89 %), *Acacia* expanded to 468 samples (11.78 %), *Acacia reficiens* substantially increased to 427 samples (10.75 %), *Maerua edulis* growing to 422 samples (10.62 %), and *Desert rose* augmented to 300 samples (7.55 %), resulting in improved plant imbalance ratio of 1:0.429, while maintaining ecological representation. Moreover, 581-image generalization test set, capturing diverse field conditions, assessed the model's robustness in real-world scenarios (Section 5).

3.2. Model architecture

Both human-based and computer vision processes for plant classification often rely on both important features like leaf shape and global context, including overall plant shape and structure. In the proposed architecture, a Swin transformer is utilized as the backbone, which is capable of storing hierarchical features. In order to improve the backbone for the extraction of plant features, we employ selective layers and transfer learning. In this model, fine-grained details, like leaf patterns, are processed along with broader features, such as the overall structure of the plant.

Fig. 2 illustrates the complete architecture of the model. The input image $I \in R^{H \times W \times 3}$ is processed by a Swin Transformer encoder, which produces hierarchical feature representations. The encoder outputs both the final encoded features $E \in R^{\frac{H}{32} \times \frac{W}{32} \times C}$ and intermediate hidden states $H = \{H_i\}_{i=1}^{4}$, where each $H_i \in R^{\frac{H}{2^{i+1}} \times \frac{W}{2^{i+1}} \times C_i}$.

3.3. Feature fusion module

The key component of our architecture is the feature fusion module, which creates rich, multiscale representations by integrating multiple levels of features from the transformer's output. The fusion process begins by extracting three types of complementary features: global context (F_{global}), local structural information (F_{local}) from the final layer, and fine-grained botanical details ($F_{fine\_grained}$) from the penultimate layer.

These features capture information at different levels of abstraction,

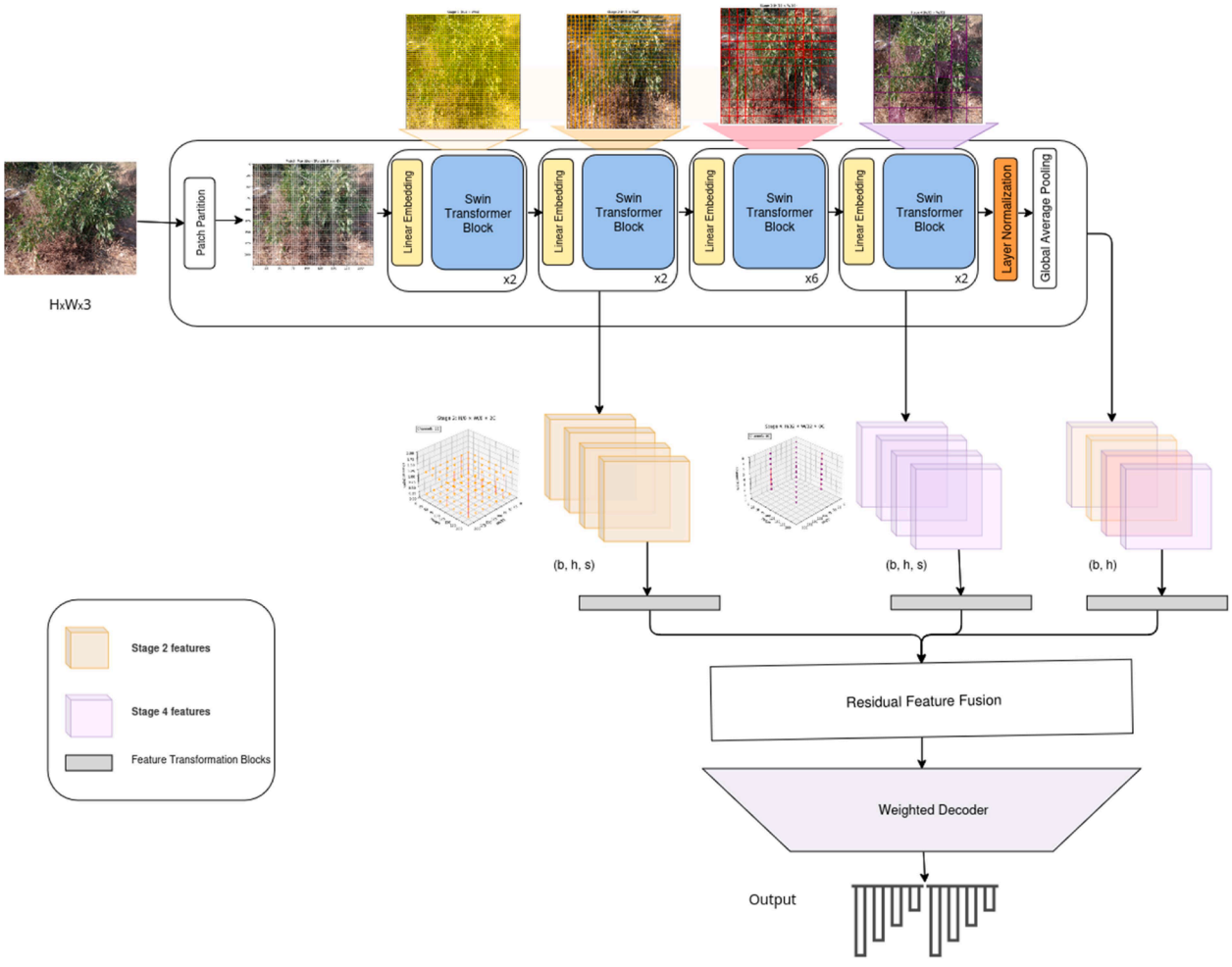

Fig. 2. Architecture of the proposed model using adaptive multiscale feature fusion.

and each feature stream is processed through a specialized projection layer [24] to enhance its representational capacity before being concatenated and fused:

$$F_{fused} = F\left(\oplus_{i\in\{global,local,fine\_grained\}} P_i(F_i)\right)$$
$$= F\left(\left[\ P_{global}\left(F_{global}\right),\ P_{local}(F_{local}),\ P_{fine\_grained}\left(F_{fine\_grained}\right)\ \right]\right)$$

where $\oplus$ denotes the concatenation operation, $P_i : R^d \to R^{d_i}$ are projection functions realized as multi-layer perceptrons with normalization and non-linear activations, and $F : R^{\sum_i d_i} \to R^d$ is the fusion function.

Moreover, in our feature fusion module, the incorporation of an adaptive residual connection mechanism is considered. This connection combines the fused representation with a weighted sum of the original features, creating a path for direct information flow that helps preserve critical botanical features during training. The multiscale feature representation is computed as:

$$F_{fused} = F(\oplus_{i\in I} P_i(F_i)) + \sum_{i\in I} \sigma_i(w)\cdot F_i$$

where $I = \left\{global, local, fine\_grained\right\}$ is the set of feature types, $\sigma(w)$ stand for softmax normalization of learnable weights initialized to prioritize global features.

In contrast to standard residual connections [22], enabling deep networks to maintain low-level image information [12,13], our approach employs learnable weights that determine the importance of each feature type and adapts during training to the specific requirements of the plant classification task. This adaptive weighting mechanism preserves feature diversity by maintaining direct paths from all abstraction levels. It allows the model to dynamically adjust the balance between global context and fine details. In addition, this mechanism facilitates the learning of complex relationships that span multiple scales. By enabling this direct path from input to output while maintaining the benefits of deep feature transformation.

3.4. Addressing class imbalance with weighted decoder

Plant databases exhibit significant class imbalance [5], with some common species appearing frequently while rare or endangered species have limited samples. To effectively handle this challenge, we implement a Weighted Decoder, which incorporates class frequency information directly into the loss function. The implemented Weighted Decoder with a class-balanced loss function, aiming to class weighting provides several advantages that allow the model to maintain high recall for rare and endangered plant species, critical for biodiversity monitoring applications.

$$L_{CE} = -\sum_{c=1}^{C} w_c \cdot y_c \log(p_c)$$

Where the class weights w_c are derived through a moderated logarithmic transformation:

$$w = \psi\left(\frac{1}{f}\right)$$

Here, ψ represents our composite moderation function that applies logarithmic scaling and range normalization to constrain weights within a controlled range, ensuring minority classes receive enhanced but not excessive weighting while preventing training instability [40].This approach facilitates a balanced learning signal throughout the class distribution thereby ensuring the accurate classification of both prevalent and rare plant species. Furthermore, the decoder incorporates label smoothing (0.1) to mitigate overconfidence and enhance generalization, an essential aspect when addressing visually similar plant species exhibiting shared taxonomic characteristics. Finally, scaling the weights to sum proportionally to the number of classes maintains proper normalization, which mitigates training instability.

4. Experiment result

In this section, the main investigation is categorized, including 3 key questions we aim to address, which are as follows:

Q1: What are the defining characteristics of the rangeland plant species image dataset?

Q2: Considering the state-of-the-art transfer learning, which is adapted for diverse tasks, how is the performance of available large models?

Q3: Can improved Multiscale fusion enhance the performance of the proposed plant classification learning model?

To address question Q1, data characteristics are investigated in Section 4.1. For Q2 (Section 4.2), different state-of-the-art models, including Swin architecture, ViT, and DeiT, are investigated for our classification task. In addition, different configurations and their effects were investigated. Finally, we applied empirical analysis to Q3 (Section 4.3) to assess different results and the effectiveness of the decoder we designed.

4.1. Q1: what are the defining characteristics of the rangeland plant species image dataset?

Fig. 3 presents a t-SNE [44] visualization of embeddings for collected rangeland plant species images in the collected dataset across arid environments. As shown, the t-SNE provides visualization demonstrating significant overlap among the plant species, indicating that high-dimensional features extracted from these plant images share considerable similarities across species. This complex distribution, where species like Yabah show only important distributional patterns insufficient for clear differentiation, highlights the challenges of plant species classification in arid environments and suggests that the strong model is required for this dataset. To address these challenges, we propose implementing intra-class augmentation techniques to each plant species' specific characteristics, including: species-specific color transformations calibrated to seasonal variations, brightness and contrast modifications simulating diverse lighting conditions in arid environments, carefully constrained geometric transformations capturing growth stage variations while preserving key morphological features, and environmental effect simulations relevant to arid regions. However, the existence of evergreen species (e.g., *Desert rose*, Prosopis juliflora) is expected to provide more clear grouping scattered in feature space, however, Fig. 3 demonstrates multiple small groupings that reflect plants changing appearance with growth stages.

Additionally, our architectural design as stated in Section 3.2, incorporates weighted mechanisms to focus on distinguishing characteristics between visually similar species and label smoothing in the cross-entropy loss function to mitigate model overconfidence when dealing with species exhibiting substantial morphological variations, all of which align with previous findings that phenological changes and environmental factors complicate plant species classification.

In collected data, *Desert rose*, known for its harmful effects on livestock based on community knowledge, exhibits a limited presence with fewer data points compared to other species. This is based on the less frequent occurrence of *Desert rose* in the arid environment studied, as evidenced by the limited number of brown points in the t-SNE visualization. However, considering the harmfulness of this species to livestock, accurate identification remains crucial. Therefore, despite its less frequent use in the dataset and underrepresentation in features, it is necessary to employ specialized machine learning techniques, such as weighted loss functions, to enhance the model's sensitivity towards *Desert rose*. As stated in 3.2 by assigning a higher weight to *Desert rose* instances during training, the model can learn to prioritize its accurate classification, even with limited training examples. This approach helps mitigate the challenges posed by the species' scarcity and ensures its

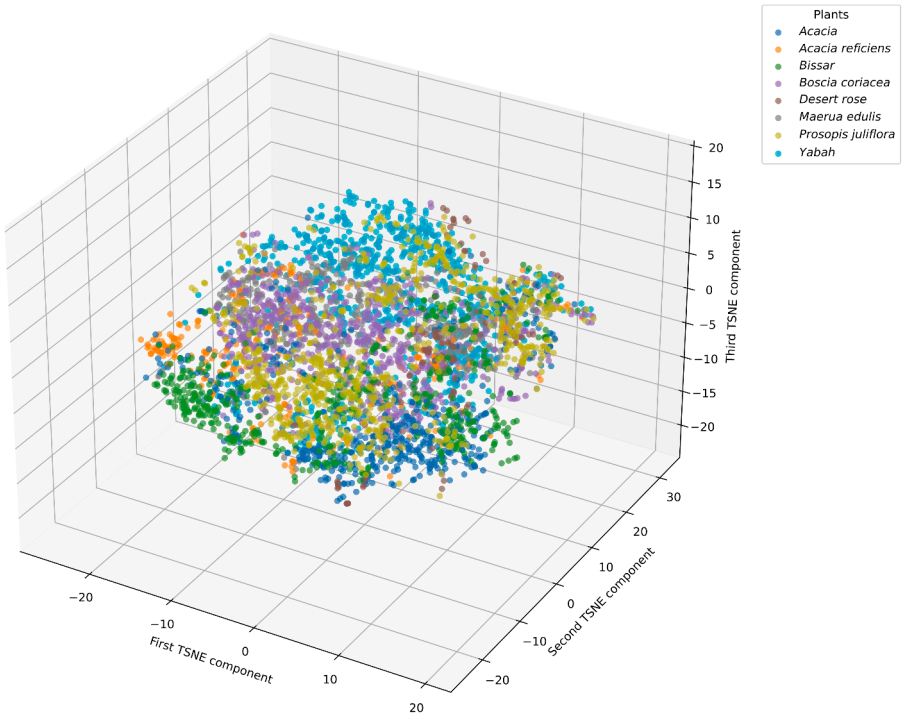

Fig. 3. T-SNE visualization of extracted features from image data.

reliable identification, which is important for livestock management and rangeland health. In contrast, the invasive *Prosopias juliflora* [7] displays a wider spread, reflecting its adaptability to various environmental conditions and growth stages. This contrast highlights the importance of understanding species-specific, especially those related to livestock health when interpreting visual patterns. Furthermore, useful species like *Acacia* and *Bissar* show even more pronounced visual changes, likely due to their responses to seasonal variations. This observation specifies the need to consider both ecological context and phenological dynamics when developing the data optimization pipeline and classification models. Consequently, this study has employed an adaptive approach to consider class with underrepresentation, addressing the complex multi-cluster patterns observed in the feature space. The transformation pipeline's diverse color adjustments, brightness/contrast modifications, and geometric transformations effectively simulate the visual variability introduced by seasonal changes.

4.2. Q2: considering the state-of-the-art transfer learning which adapted for diverse tasks how the performance of available large models is?

To investigate state-of-the-art transfer learning, this study has selected three transformer-based models to assess the transferability of extracted features from plant-based data. This study examines transfer learning performance on rangeland-based species are using by herders using three transformer-based models: ViT (Vision Transformer) [11], DeiT [42] (Data-efficient Image Transformer), and Swin [34]. Our experiments were conducted using the Google Colaboratory Pro+ platform, which provided computational resources including an NVIDIA Tesla T4 GPU with 15.0 GB of GPU memory and 51.0 GB of system RAM. This infrastructure enabled efficient model training and evaluation processes. The development environment was built on Python 3, with PyTorch as the primary deep learning framework for implementing and training our neural network architecture. The models were trained using a learning rate of $2e-5$, training was conducted for 6 epochs with a batch size of 64 per device.

Based on the comparative analysis in Table 1, the Swin Transformer architecture using a Balanced Decoder (BD) represents acceptable performance t in how vision transformer models handle varying window sizes and resolutions. Swin-*B*+BD and Swin-*B*+WD demonstrate better performance with consistently high metrics across various species categories, maintaining higher precision values as recall increases. The performance differences between these models are based on their architectural distinctions. Swin Transformer's hierarchical design with

Table 1
Comparative performance of vision transformer-based Ppant species classification.

Model	Plant Species	Prec.	Rec.	F1
ViT-*B*+BD	Yabah	79.6 %	80.4 %	80.0 %
	Prosopis juliflora	89.0 %	83.5 %	86.1 %
	Acacia	69.7 %	73.1 %	71.4 %
	Bissar	89.3 %	84.7 %	86.9 %
	Desert rose	81.0 %	65.4 %	72.3 %
	Boscia coriacea	75.0 %	74.4 %	74.7 %
	Maerua edulis	77.3 %	77.3 %	77.3 %
	Acacia reficiens	50.0 %	61.8 %	55.3 %
	Average	76.3 %	75.1 %	75.5 %
	Overall Model Acc. : 77.5 %			
DeiT-*B*+BD	Yabah	79.7 %	83.2 %	81.4 %
	Prosopis juliflora	85.0 %	77.9 %	81.3 %
	Acacia	66.4 %	72.1 %	69.1 %
	Bissar	89.4 %	85.7 %	87.5 %
	Desert rose	87.5 %	80.8 %	84.0 %
	Boscia coriacea	72.7 %	63.9 %	68.0 %
	Maerua edulis	71.8 %	80.9 %	76.1 %
	Acacia reficiens	58.6 %	61.8 %	60.2 %
	Average	76.4 %	75.8 %	75.9 %
	Overall Model Acc. : 77.5 %			
Swin-*B*+BD	Yabah	89.0 %	91.9 %	90.4 %
	Prosopis juliflora	92.4 %	92.4 %	92.4 %
	Acacia	84.5 %	78.9 %	81.6 %
	Bissar	91.2 %	94.9 %	93.0 %
	Desert rose	96.0 %	92.3 %	94.1 %
	Boscia coriacea	86.9 %	85.0 %	85.9 %
	Maerua edulis	87.3 %	87.3 %	87.3 %
	Acacia reficiens	73.2 %	74.6 %	73.9 %
	Average	87.6 %	87.1 %	87.3 %
	Overall Model Acc. : 88.0 %			

shifted windows allows for both local feature extraction and global context modeling while maintaining computational efficiency. This architecture effectively addresses the standard vision transformer's limitations by introducing locality into the self-attention mechanism and adapting to different image scales, a critical advantage when classifying plants photographed from inconsistent distances in field conditions. Moreover, Fig. 4 demonstrates image-classification metrics. Swin-*B*+BD (4.a) demonstrates more consistent performance across all plant species in both the bar charts and precision-recall curves compared to ViT (Fig. 4.b) and DeiT (Fig. 4.c). While models like ViT compute attention globally and DeiT maintains similar constraints, Swin-*B*+BD utilizes a flexible local window approach that adapts during feature extraction, allowing for more effective feature recognition across different tree species. The stated result also supports t-SNE visualization in Fig. 3, not only demonstrating the dataset's structure but also supporting the choice of the Swin Transformer as a suitable architecture using a hierarchical shifted-window approach enabling both local and global attention for the image-based plant species with a range of variety.

4.3. Q3: can improved multiscale fusion enhance the performance of the proposed plant classification learning model?

This section aims to evaluate the proposed Multiscale approach and fusing features from different scales that can increase the performance of plant classification. Considering Table 2 and Table 1, the Swin transformer performs well. The base Swin-*B*+BD model employs a Swin Transformer base variant combined with the BD described in Section 4.2.

The proposed Multiscale enhances this architecture by implementing the feature fusion method, which combines global features with fine-grained features from earlier layers, creating a stronger representation that captures both overall plant structure and specific details like leaf patterns. The Multiscale approach, using a weighted decoder, outperformed the Swin base model, achieving a 89.71 % accuracy,

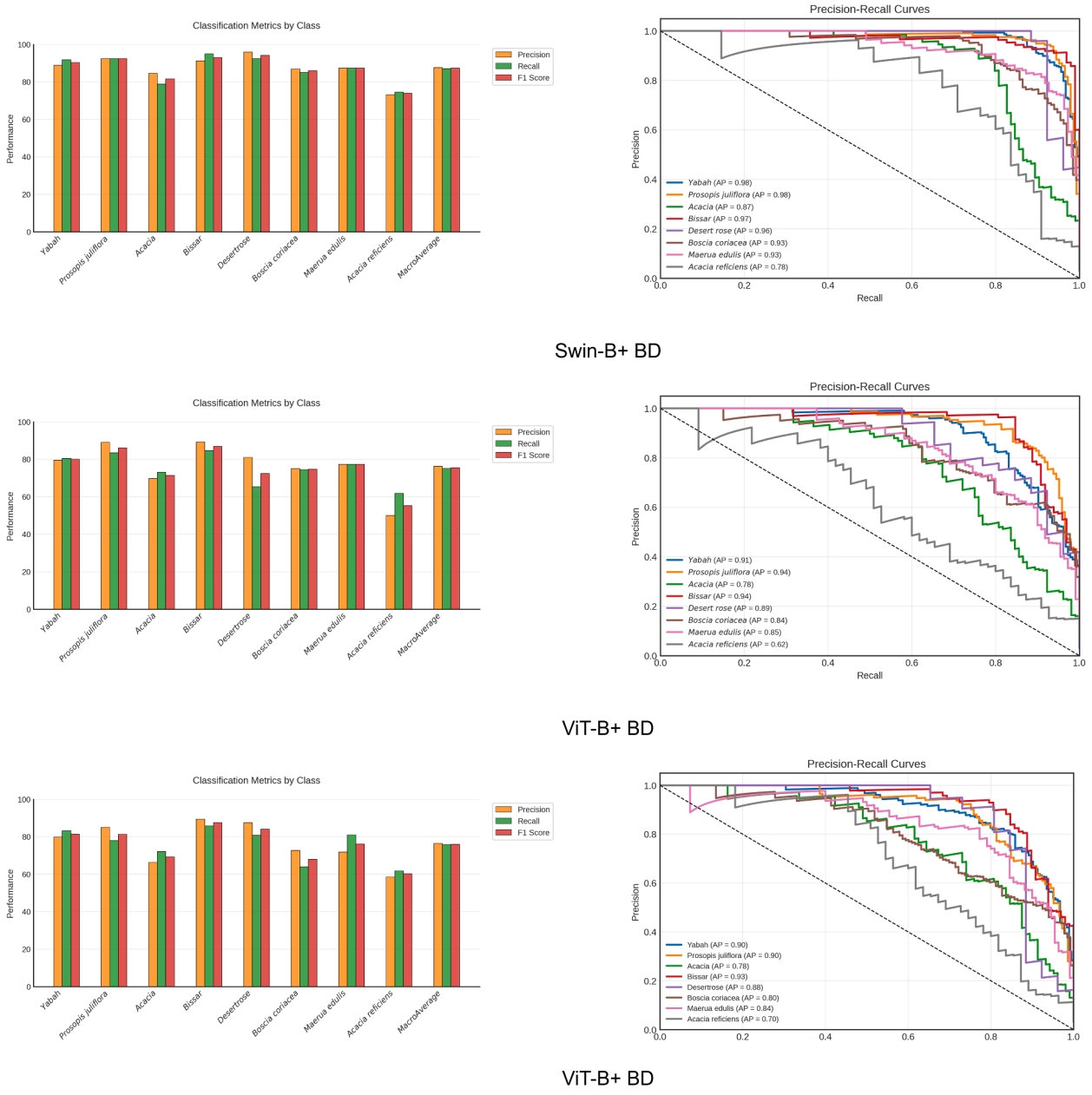

Fig. 4. Evaluation of multi-class plant species classification models, including the classification metrics (left) and precision-recall curves (right) for performance comparison.

Table 2
Experimental result and comparison proposed Multiscale fusion Swin model.

	Multiscale EfficientNet			Multiscale Swin-*B*+WD		
Class	Prec.	Rec.	F1	Prec.	Rec.	F1
Yabah	87.13 %	80.98 %	83.94 %	95.56 %	93.48 %	94.51 %
Prosopis juliflora	85.31 %	84.14 %	84.72 %	97.06 %	91.03 %	93.95 %
Acacia	65.00 %	75.00 %	69.64 %	77.98 %	81.73 %	79.81 %
Bissar	85.05 %	92.86 %	88.78 %	92.16 %	95.92 %	94.00 %
Desert rose	84.62 %	84.62 %	84.62 %	88.89 %	92.31 %	90.57 %
Boscia coriacea	68.24 %	75.94 %	71.89 %	86.11 %	93.23 %	89.53 %
Maerua edulis	76.53 %	68.18 %	72.12 %	90.48 %	86.36 %	88.37 %
Acacia reficiens	80.95 %	61.82 %	70.10 %	78.85 %	74.55 %	76.64 %
Average	79.10 %	77.94 %	78.23 %	88.38 %	88.58 %	88.42 %
	Overall Model Acc. : 78.60 %			Overall Model Acc. : 89.71 %		

demonstrating that incorporating features at different scales helps the model better recognize and classify plant species by leveraging both local details and global context. Moreover, Fig. 5 demonstrates the proposed Multiscale model's performance across all plant species, with Yabah, *Prosopis juliflora*, and *Bissar* achieving promising performance, considering Average Precision (AP) is the metric that measures the area under the precision-recall curve, and higher AP means the model can identify plant species with minimal errors in real-world practice, which is important for botanical surveys, invasive species monitoring, and ecological research where misidentification could lead to incorrect future tasks like pasture scoring and rangeland management. Considering class individual performance, For *Desert rose*, while both models exhibit robust results. More integrated improvements are observed for *Boscia coriacea* respectively, 89.53 % F1 and 93.23 % and *Maerua edulis* 88.37 % F1 versus 87.3 % F1, indicating the Multiscale approach's usefulness in these categories. Notably, the Multiscale model enhances the classification of challenging species like *Acacia reficiens*, which, the F1 score improves from 73.9 % to 76.64 %. These improvements suggest the Multiscale approach's ability to capture both fine-grained leaf patterns and broader structural context, for distinguishing these species. The performance improvements are in the precision-recall curves, which demonstrate higher precision across varying recall levels. Moreover, our dataset demonstrates common class imbalance patterns found in biodiversity collections, with sample distributions ranging from 3 % for underrepresented species to over 21% samples for Yabah as majority class.

To address this challenge, the weighted decoder employs an approach that transforms class frequency information through inverse frequency weighting, logarithmic moderation, and controlled range normalization to maintain weights. This approach creates differentiated learning dynamics where weight assignments different ranging for highly and underrepresented species produce varying gradient amplification during backpropagation, improving minority class features while preventing majority class dominance. The experimental results demonstrate an overall accuracy of 89.71 % with an average recall of 88.58 %, indicating that the weighted loss function maintains competitive classification performance across taxonomic groups. Performance variations demonstrated with well-represented species achieving high classification metrics (e.g., Yabah 93.48 %, *Prosopis juliflora* 91.03 %) while species show recall-favored profiles (e.g., *Desert rose* 92.31 % recall vs 88.89 % precision), this could suggest successful gradient amplification for underrepresented classes. As demonstrated in the result, precision range of 77.98% to 97.06 % , with no species showing degraded performance, demonstrate that the weight moderation approach maintained reasonable bounds all classes. The similar performance for morphologically related species (*Acacia* 79.81% vs *Acacia reficiens* 76.64%) indicates that intra-genus classification challenges may persist despite frequency-based reweighting, while the consistent high recall rates suggest compatibility with applications requiring comprehensive taxonomic coverage in biodiversity assessment scenarios.

To further validate the effectiveness of the proposed transformer-based multiscale approach compared with CNN-Based models, we chose EfficientNet-B4 [55] as CNN baseline, aiming for a model comparable in complexity to the Swin-B architecture. The comparative analysis as demonstrated in Table 2, reveals the proposed Multiscale Swin-*B*+WD model outperforms the CNN-based EfficientNet approach throughout evaluation metrics. The Multiscale Swin model achieves an overall model accuracy of 89.71 % compared to EfficientNet's 78.60 %. These results demonstrate that the attention-based mechanisms in transformer architectures, combined with the multiscale feature fusion approach, provide enhanced capability for capturing both local botanical features and global structural patterns essential for accurate plant species classification.

Fig. 6 shows the Feature Attribution and Gradient-weighted Class Activation Mapping (GradCAM). The visualizations show that the model focuses on plant features like crown morphology and growth habit, which aligns with established botanical identification practices. These visualizations offer valuable insights into the spatial localization of influential features within a model's input.

5. Discussion

Herders rely on rangeland ecosystems to sustain their livestock.

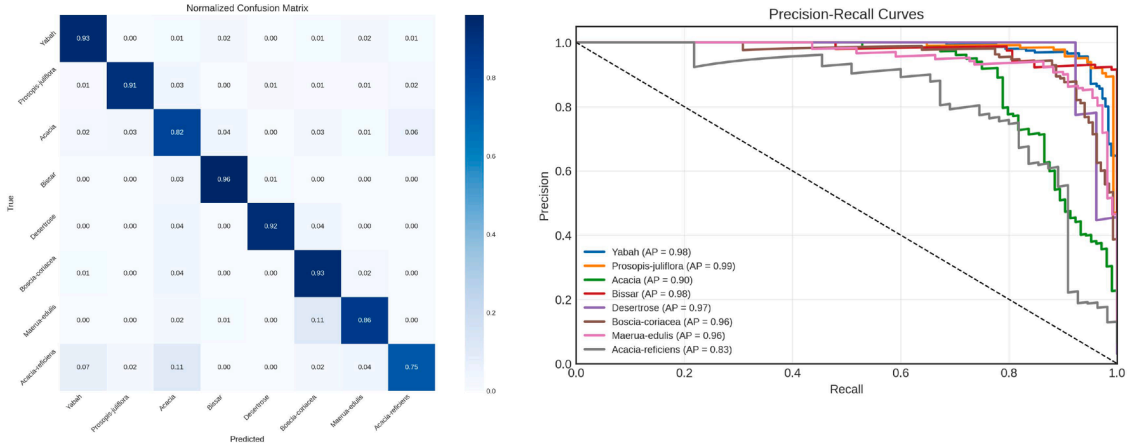

Fig. 5. The performance of proposed Multiscale fusion demonstrated in (a) Confusion matrices and (b) Precision-recall curves.

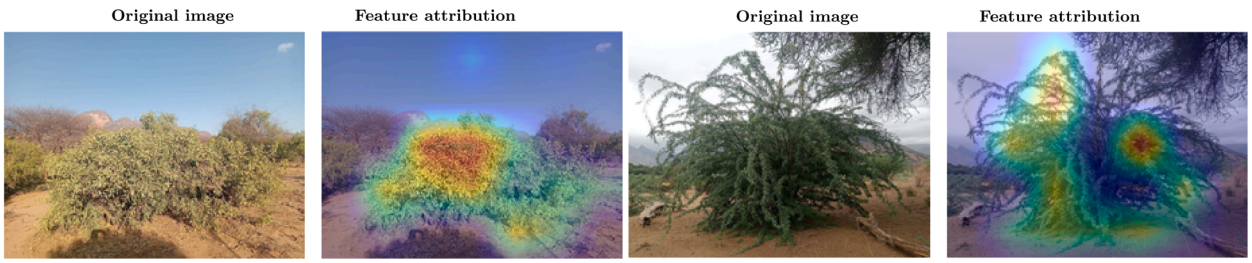

Fig. 6. Correct classification of *Boscia coriacea* (left) *Prosopis juliflora* (right) considering different growth stages and environments.

Plant-based assessing pasture quality is important for making informed management decisions. In this context, one of the main tasks of rangeland-based livestock feeding is to extract as many as possible useful vegetation indicators. To address this need, this paper introduced a novel Multiscale plant classification learning model, aiming to achieve a more comprehensive extraction of information for plant image analysis. Among studies on plant classification across diverse domains, CNN-based models have been predominantly employed, demonstrating promising performance, particularly in controlled environments. However, the proposed multiscale approach addresses plant classification within the heterogeneous natural environment of rangelands, focusing on broader botanical categories relevant to livestock feeding, such as trees, shrubs, herbs, and grasses. Given the unique characteristics of our self-collected dataset and the relative scarcity of research specifically addressing plant-based livestock feeding in such environments, we conducted a comparative analysis, presented in Table 1 and Table 2, evaluating the performance of three established transformer-based baseline models comparing the proposed multiscale approach. As demonstrated on table F1-scores, which balance precision and recall, are also high for different plant types comparing model results. The important discussion based on our findings are:

- **Data collection challenges**: Several helpful interactive photographic guides have been introduced for African plant species, such as "African Plants a Photo Guide" [21] . These guides, while helpful for human identification and education, fall short of the requirements for building dataset for machine learning models. Apart from copyright issues, they typically lack the required volume of images needed to train an algorithm to recognize the nuances and variations within a species. Furthermore, even for Fewshot learning, the images within these guides are often visually harmonious and clear, which means they might not capture the full range of real-world conditions an algorithm would encounter. For instance, a single plant species in a photo guide might be shown in perfect lighting and from a single angle, whereas a machine learning dataset would require images of that species in varying stages of growth under different lighting conditions. In this work, although the mobile-based data collection facilitates this process, it was required to communicate with different local researchers and stockholders like herders to collect data in different pastures in rangeland, which are less accessible.
- **Non-constant appearance and Sparse rare plant species**: The data investigation demonstrates that sparse vegetation and inter-species variations in arid environments present considerable challenges for image-based plant identification. Considering the first question Section 4.1, although the existence of evergreen expected constant appearance to provides more structured pattern but species like *Prosopiasjuliflora* [7] displays a wider spread, reflecting its adaptability to various environmental conditions and growth stages. This contrast highlights the importance of understanding species-specific characteristics, especially those related to livestock health, when interpreting visual patterns. Furthermore, useful species like *Acacia* and *Bissar* show even more visual changes, likely due to their responses to seasonal variations. This observation highlighted the need to consider both ecological context and phenological dynamics when developing the data optimization pipeline and classification models. Furthermore, considering the natural environment species distribution, while data investigation shows three herbs and invasive species are common majority classes. In contrast, minority classes in the dataset fall into two categories: first, rare species (e.g., poisonous) plants that have a sparse distribution in the rangeland environment, and second, species that grow in less accessible areas (e.g., mountains), which require more time and cost for collecting these species samples. Moreover, there are many research opportunities for filling this gap using synthetic image generation.
- **Impact of combining the Local-Global features in plant image analysis**: Considering Q3 4.3, as demonstrated in Table 2, the obtained results on Q2 4.2 demonstrate the performance of the hierarchical architecture, furthermore, the results show that the proposed model, with its ability to connect and combine information from different levels of detail, proves for accurate species identification. Considering, Swin extracted features in early stages that capture more fine-grained, local details and texture information. These features of the information have a higher spatial resolution, allowing it to represent smaller plant details and finer details better. On the other side, in later stages, the spatial resolution decreases, and the Swin network focuses on capturing more global, semantic, and abstract information. To this end, by integrating features from both fine-grained textures (e.g., leaf and branch patterns) and broader structural elements (e.g., plant morphology), the model gains a more complete understanding of each species in our dataset, which has variety in growing stage and seasonal timing. This multi-scale analysis generally obtained better performance among the compared strategies 1, especially for similar species Fig. 6, which is beneficial for distinguishing challenging species like, where implicit differences that might be overlooked when considering only isolated features become clearer. The improved performance AP metric for such species demonstrates the advantage of this approach.
- **Generalizability**: In this study different machine learning models have been investigated, ViT, Swin, and DeiT, in general, utilize similar foundational datasets, which ImageNet [9] is a well-known diverse dataset for computer vision, the results demonstrate that these models trained on this public dataset, despite the fewer data from Africa [10] can provide acceptable performance and general ability on our plant-based dataset which has a natural-based imbalance distribution.

Moreover, to evaluate the model's generalization capabilities across more diverse field conditions, a dedicated generalization test set was provided. This set includes 581 plant image samples collected from different environments or times, demonstrating variations in background, and plant growth stages. The test set aimed to assess the model's ability to maintain performance despite the heterogeneity of natural rangeland conditions. This independent evaluation provided a measure of the model's robustness and its suitability for practical deployment in

real-world applications.

Table 3 presents a performance comparison of three transformer-based models, including our proposed Multiscale Swin-*B*+WD, in plant species classification on a generalization test set. The results demonstrate the promising performance of the Multiscale Swin-B using weighted decoder (WD) model, outperforming the other models across most evaluation metrics. It achieves an average accuracy evaluation metric of precision of 78.80 %, a recall of 73.67 %, and an F1-score of 74.43 %. In comparison, the Swin-*B*+BD model achieves an average Recall of 72.12 % and an F1-score of 71.99 %, while the ViT-*B*+BD model shows the lowest performance. with an average metrics of 64.54 % and an F1-score of 64.65 %. The Multiscale Swin-*B*+WD, in contrast ViT, employs both hierarchical structures with shifted windows (Swin-B) and multi-scale fusion, enabling it to model fine-grained details and long-range dependencies, important for distinguishing between visually similar plant species, leading to improved accuracy compared to base-models. For example, Fig. 6 illustrates the successful classification of Prosopis juliflora and *Boscia coriacea* across various growth stages and environments, highlighting the model's focus on leaf and branch patterns. However, The result demonstrate poor performance in some species like *Acacia reficiens* [21], which as demonstrated in the plant photo guide, there is significant morphology difference from a shrub-like early stage to a mature tree form presents a challenge for computer vision models trained predominantly on one growth phase; consequently, all model trained on shrub-form specimens have reduced performance when applied to environments where these species are predominantly present in their mature tree morphology due to significant differences in visual features across life stages. Additionally, the results in Table 4 show differences in confidence calibration properties across models. Multiscale Swin-*B*+WD maintains higher evaluation metric like 74.35 % at base threshold and 83.45 % at >0.5 while preserving almost 73.0 % of samples. In comparison, Swin-*B*+BD reaches 91.42 % performance but retains fewer samples (58.2 %). In the context of herd movement, reliable results for plant identification, built upon a foundation of accurate and comprehensive data, act as a powerful information tool, which is essential for real-world applications in plant identification and monitoring where reliability is important in the context of decision-making systems (e.g., herd movement) and the empowerment of diverse user groups (e.g., local communities) in real-world applications.

Table 3
Performance comparison classifier using base transformer models and proposed model on the generalization test set. The model achieved better performance, yet all models exhibited limitations in classifying species with significant morphology difference from a shrub-like early stage to a mature tree in different environment (e.g., Acacia reficiens).

Model	Plant Species	Prec.	Rec.	F1	Total Samples
ViT-*B*+BD	Yabah	83.43 %	71.57 %	77.05 %	197
	Prosopis juliflora	63.70 %	83.04 %	72.09 %	112
	Acacia	62.50 %	71.43 %	66.67 %	77
	Bissar	69.77 %	48.39 %	57.14 %	62
	Desert rose	23.68 %	75.00 %	36.00 %	12
	Boscia coriacea	45.07 %	39.51 %	42.11 %	81
	Maerua edulis	66.67 %	50.00 %	57.14 %	20
	Acacia reficiens	45.45 %	25.00 %	32.26 %	20
	Average	66.93 %	64.54 %	64.65 %	581
Swin-*B*+BD	Yabah	83.07 %	79.70 %	81.35 %	197
	Prosopis juliflora	86.84 %	88.39 %	87.61 %	112
	Acacia	63.11 %	84.42 %	72.22 %	77
	Bissar	81.63 %	64.52 %	72.07 %	62
	Desert rose	37.50 %	50.00 %	42.86 %	12
	Boscia coriacea	45.68 %	45.68 %	45.68 %	81
	Maerua edulis	64.29 %	45.00 %	52.94 %	20
	Acacia reficiens	40.00 %	30.00 %	34.29 %	20
	Average	72.71 %	72.12 %	71.99 %	581
Multiscale Swin-*B*+WD	Yabah	95.56 %	65.48 %	77.71 %	197
	Prosopis juliflora	92.66 %	90.18 %	91.40 %	112
	Acacia	62.16 %	89.61 %	73.40 %	77
	Bissar	88.14 %	83.87 %	85.95 %	62
	Desert rose	30.30 %	83.33 %	44.44 %	12
	Boscia coriacea	48.84 %	51.85 %	50.30 %	81
	Maerua edulis	54.29 %	95.00 %	69.09 %	20
	Acacia reficiens	46.15 %	30.00 %	36.36 %	20
	Average	**78.80 %**	**73.67 %**	**74.43 %**	**581**

6. Conclusions

This study evaluated the performance of transformer architectures for rangeland-based plant species classification, proposing a Multiscale model that demonstrably enhances image classification through architectural design and feature fusion. The result of this study can be understood as evidence that transformer architectures and transfer learning, trained on public datasets, show promising performance in plant species classification tasks, offering a promising solution for challenges posed by data scarcity and spatial sparsity, which are common in African plant data. On this basis, we conclude that the proposed Multiscale approach with a weighted decoder demonstrates promising performance, achieving 89.71 % accuracy and outperforming the Swin

Table 4
Performance comparison classifier results and related Confidence threshold for classifying the generalization test set. The Multiscale model shows an acceptable balance, maintaining classification metrics for a larger portion of samples, while the other model achieves higher performance on fewer samples.

Model	Prec.	Rec.	F1	Conf. (threshold)	Retained Samples
ViT-*B*+BD	66.93 %	64.54 %	64.65 %	> 0.0	581
	69.43 %	67.68 %	67.66 %	> 0.3	526
	86.97 %	85.87 %	85.57 %	> 0.5	**269**
Swin-*B*+BD	72.71 %	72.12 %	71.99 %	> 0.0	581
	78.74 %	77.44 %	77.53 %	> 0.3	523
	92.95 %	91.42%	91.45 %	> 0.5	**338**
Multiscale Swin-*B*+WD	78.84 %	74.35 %	74.77 %	> 0.0	581
	79.66 %	75.97 %	76.23 %	> 0.3	566
	86.30 %	**83.45 %**	**83.33 %**	> 0.5	**435**

base model's 88.0 % accuracy. Moreover, our study suggests that while some herders do utilize rangeland images for personal assessment, this data is often fragmented and stored to their individual devices, limiting its broader utility and scalability. To this end, the proposed model can provide automated bulk rangeland species image classification, allowing for more precise pasture scoring, which is vital for effective and sustainable grazing management when traditional seasonal patterns are disrupted by climate variability. This conclusion follows from the fact that the integration of features derived from fine-grained textural information, such as leaf and branch patterns, with broader structural characteristics, including plant morphology, contributes to a more comprehensive species representation. However, it is crucial to acknowledge the broader ecological context, specifically the significant relationship between global warming and land use systems such as pastoralism, which utilize rangeland ecosystems for grazing. As climate change impacts these ecosystems and alters species distributions, future research should consider the implications for plant species classification. Future research should be devoted to the development of expanded data collection and the exploration of emerging large-scale models, as well as the integration of ecological factors into data optimization strategies using image generation techniques. This integrated approach will ensure the adaptability and effectiveness of classification models in the face of evolving environmental conditions.

Ethics statement

Not applicable: This manuscript does not include human or animal research.

If this manuscript involves research on animals or humans, it is imperative to disclose all approval details.

If Yes, please provide your text here:

CRediT authorship contribution statement

Zakieh Alizadehsani: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation. **Oliver Hensel:** Writing – review & editing, Supervision, Resources, Funding acquisition. **Abozar Nasirahmadi:** Writing – review & editing, Supervision, Project administration, Conceptualization.

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.

Acknowledgments

This research was conducted as part of the InfoRange project, funded by the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF) with the grant number "01LL2201B". We sincerely thank the funding bodies for supporting this research. We would also like to express our gratitude to Compwiz Creations (https://compwiz.io/) for their support in developing the app used for part of the data collection. Additionally, we appreciate the contributions of our partners in Kenya from the InfoRange project for their support in image data collection.

Data availability

Data will be made available on request.

References

[1] Jason Adams, Yumou Qiu, Yuhang Xu, James C Schnable, Plant segmentation by supervised machine learning methods, Plant Phenome J. 3 (1) (2020) e20001.
[2] Olalekan Adekola, Jessica Lamond, Ibidun Adelekan, Namrata Bhattacharya-Mis, Mboto Ekinya, Eze Eze Bassey, Ujoh Fanan, Towards Adoption of Mobile Data Collection for Effective Adaptation and Climate Risk Management in Africa, Geosci. Data J. 10 (2) (2023) 276–290.
[3] African Plants a Photo Guide, Africanplants (2014). https://www.africanplants.senckenberg.de/root/index.php.
[4] Rahim Azadnia, Faramarz Noei-Khodabadi, Azad Moloudzadeh, Ahmad Jahanbakhshi, Mahmoud Omid, Medicinal and Poisonous Plants Classification from Visual Characteristics of Leaves Using Computer Vision and Deep Neural Networks, Ecol. Inf. 82 (2024) 102683.
[5] Jayme Garcia Arnal Barbedo, Impact of Dataset Size and Variety on the Effectiveness of Deep Learning and Transfer Learning for Plant Disease Classification, Comput. Electron. Agric. 153 (2018) 46–53.
[6] Vinit Bodhwani, D.P. Acharjya, Umesh Bodhwani, Deep Residual Networks for Plant Identification, Procedia Comput. Sci. 152 (2019) 186–194.
[7] Simon Choge, Purity Rima Mbaabu, Gabriel Mukuria Muturi, Management and Control of the Invasive Prosopis Juliflora Tree Species in Africa with a Focus on Kenya. Prosopis As a Heat Tolerant Nitrogen Fixing Desert Food Legume, Elsevier, 2022, pp. 67–81.
[8] James Clarke, Sarah Barman, Paolo Remagnino, Ken Bailey, Don Kirkup, Simon Mayo, Paul Wilkin, Venation Pattern Analysis of Leaf Images, in: Advances in Visual Computing: Second International Symposium, ISVC 2006 Lake Tahoe, NV, USA, November 6-8, 2006. Proceedings, Part II 2, Springer, 2006, pp. 427–436.
[9] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, Li Fei-Fei, Imagenet: a Large-Scale Hierarchical Image Database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, Ieee, 2009, pp. 248–255.
[10] Victor. Dibia, Cocoafrica: a Curation Tool and Dataset of Common Objects in the Context of Africa. 2nd Black in AI Workshop, NeurIPS, 2018.
[11] Dosovitskiy, Alexey, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, et al. 2020."An Image Is Worth 16x16 Words: transformers for Image Recognition at Scale." *arXiv Preprint arXiv:2010.11929*.
[12] Mads Dyrmann, Henrik Karstoft, Henrik Skov Midtiby, Plant Species Classification Using Deep Convolutional Neural Network, Biosyst. Eng. 151 (2016) 72–80.
[13] Xiangpeng Fan, Tan Sun, Xiujuan Chai, Jianping Zhou, YOLO-WDNet: a Lightweight and Accurate Model for Weeds Detection in Cotton Field, Comput. Electron. Agric. 225 (2024) 109317.
[14] Steven Franzel, Sammy Carsan, Ben Lukuyu, Judith Sinja, Charles Wambugu, Fodder Trees for Improving Livestock Productivity and Smallholder Livelihoods in Africa, Curr Opin Env. Sustain. 6 (2014) 98–103.
[15] G. Geetharamani, Arun Pandian, Identification of Plant Leaf Diseases Using a Nine-Layer Deep Convolutional Neural Network, Comput. Electr. Eng. 76 (2019) 323–338.
[16] Mostafa Mehdipour Ghazi, Berrin Yanikoglu, Erchan Aptoula, Plant Identification Using Deep Neural Networks via Optimization of Transfer Learning Parameters, Neurocomputing 235 (2017) 228–235.
[17] Dario Gogoll, Philipp Lottes, Jan Weyler, Unsupervised Domain Adaptation for Transferring Plant Classification Systems to New Field Environments, Crops, and Robots, in: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2020, pp. 2636–2642.
[18] Pushkar Gole, Punam Bedi, Sudeep Marwaha, Md Ashraful Haque, Chandan Kumar Deb, TrIncNet: a Lightweight Vision Transformer Network for Identification of Plant Diseases, Front Plant Sci. 14 (2023) 1221557.
[19] Cristina Gouveia, Alexandra Fonseca, António Câmara, Francisco Ferreira, Promoting the Use of Environmental Data Collected by Concerned Citizens Through Information and Communication Technologies, J. Env. Manage. 71 (2) (2004) 135–154.
[20] Yifan Guo, Yanting Lan, Xiaodong Chen, CST: convolutional Swin Transformer for Detecting the Degree and Types of Plant Diseases, Comput. Electron. Agric. 202 (2022) 107407.
[21] Reginald T Guuroh, Leslie R Brown, Miguel Alvarez, Manfred Finckh, Ute Schmiedel, Gaolathe Tsheboeng, Jürgen Dengler, African Vegetation Studies: introduction to a Special Collection. Vegetation Classification and Survey, Pensoft Publishers, 2024.
[22] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, Deep Residual Learning for Image Recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[23] David J. Hearn, Shape Analysis for the Automated Identification of Plants from Images of Leaves, Taxon 58 (3) (2009) 934–954.
[24] Hinton, Geoffrey, Oriol Vinyals, and Jeff Dean. 2015."Distilling the Knowledge in a Neural Network." *arXiv Preprint arXiv:1503.02531*.
[25] Md Abrar Istiak, Razib Hayat Khan, Jahid Hasan Rony, M.M. Mahbubul Syeed, M. Ashrafuzzaman, Md Rajaul Karim, Md Shakhawat Hossain, Mohammad Faisal Uddin, AqUavplant Dataset: a High-Resolution Aquatic Plant Classification and Segmentation Image Dataset Using UAV, Sci. Data 11 (1) (2024) 1411.
[26] Ponugoti Kalpana, R. Anandan, Abdelazim G Hussien, Hazem Migdady, Laith Abualigah, Plant Disease Recognition Using Residual Convolutional Enlightened Swin Transformer Networks, Sci. Rep. 14 (1) (2024) 8660.
[27] J.-K. Kamarainen, Ville Kyrki, Heikki Kalviainen, Invariance Properties of Gabor Filter-Based Features-Overview and Applications, IEEE Trans. Image Process. 15 (5) (2006) 1088–1099.

[28] Kaur, Rajwant, Wilfredo L Gonzales, Luis Daniel Llambi, Pascual J Soriano, Ragan M Callaway, Marnie E Rout, Timothy J Gallaher, and Inderjit. 2012."Community Impacts of Prosopis Juliflora Invasion: biogeographic and Congeneric Comparisons".
[29] Soon Jye Kho, Sugumaran Manickam, Sorayya Malek, Mogeeb Mosleh, Sarinder Kaur Dhillon, Automated Plant Identification Using Artificial Neural Network and Support Vector Machine, Front Life Sci. 10 (1) (2017) 98–107.
[30] Alex Krizhevsky, Ilya Sutskever, Geoffrey E Hinton, Imagenet Classification with Deep Convolutional Neural Networks, Adv. Neural Inf. Proc. Syst. (2012) 25.
[31] Neeraj Kumar, Peter N Belhumeur, Arijit Biswas, David W Jacobs, W. John Kress, Ida C Lopez, João VB Soares, Leafsnap: a Computer Vision System for Automatic Plant Species Identification, in: Computer Vision–ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part II 12, Springer, 2012, pp. 502–516.
[32] Mónica G Larese, Rafael Namías, Roque M Craviotto, Miriam R Arango, Carina Gallo, Pablo M Granitto, Automatic Classification of Legumes Using Leaf Vein Image Features, Pattern Recognit. 47 (1) (2014) 158–168.
[33] Sue Han Lee, Chee Seng Chan, Paul Wilkin, Paolo Remagnino, Deep-Plant: plant Identification with Convolutional Neural Networks, in: 2015 IEEE International Conference on Image Processing (ICIP), IEEE, 2015, pp. 452–456.
[34] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, Baining Guo, Swin Transformer: hierarchical Vision Transformer Using Shifted Windows, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 10012–10022.
[35] João Camargo Neto, George E Meyer, David D Jones, Ashok K Samal, Plant Species Identification Using Elliptic Fourier Leaf Shape Analysis, Comput. Electron. Agric. 50 (2) (2006) 121–134.
[36] J.M. Rust, T. Rust, Climate Change and Livestock Production: a Review with Emphasis on Africa, S. Afr. J. Anim. Sci. 43 (3) (2013) 255–267.
[37] Scott A Shearer, R.G. Holmes, Plant Identification Using Color Co-Occurrence Matrices, Trans. ASAE 33 (6) (1990) 1237–1244.
[38] Simonyan, Karen, and Andrew Zisserman. 2014."Very Deep Convolutional Networks for Large-Scale Image Recognition." *arXiv Preprint arXiv:1409.1556*.
[39] Yu Sun, Yuan Liu, Guan Wang, Haiyan Zhang, Deep Learning for Plant Identification in Natural Environment, Comput. Intell. Neurosci. 2017 (1) (2017) 7361042.
[40] Z. Lai, C. Wang, H. Gunawan, S.C.S. Cheung, C.N. Chuah, Smoothed adaptive weighting for imbalanced semi-supervised learning: improve reliability against unknown distribution data, in: International Conference on Machine Learning, PMLR, 2022, June, pp. 11828–11843.
[41] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich, Going Deeper with Convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.
[42] Hugo Touvron, Matthieu Cord, Matthijs Douze, Francisco Massa, Alexandre Sablayrolles, Hervé Jégou, Training Data-Efficient Image Transformers & Distillation Through Attention, in: International Conference on Machine Learning, PMLR, 2021, pp. 10347–10357.
[43] Turnbull, Sophie, and Connor Harrison. 2024."Shepherd's Eye In The Sky: The Potential For Afriscout Digital Grazing Maps To Improve Pastoralists'grazing And Migration Decisions".
[44] Laurens Van der Maaten, Geoffrey Hinton, Visualizing Data Using t-SNE, J. Mach. Learn. Res. 9 (11) (2008).
[45] Shivali Amit Wagle, et al., Comparison of Plant Leaf Classification Using Modified AlexNet and Support Vector Machine, Trait. Du Signal 38 (1) (2021).
[46] Xuan Wang, Junhua Liang, Fangxia Guo, Feature Extraction Algorithm Based on Dual-Scale Decomposition and Local Binary Descriptors for Plant Leaf Recognition, Digit Signal Proc. 34 (2014) 101–107.
[47] O. Wasonga, John Musembi, Kennedy Rotich, Ibrahim Jarso, Caroline King-Okumu, R.K. Kyuma, et al., Vegetation Resources and Their Economic Importance in Isiolo County, Kenya, in: International Institute for Environment and Development (IIED), London, 2016. https://Pubs.Iied.Org/10141IIED.
[48] Wei Tan, Siow-Wee Jing, Sameem Chang, Hwa Abdul-Kareem, Jen Yap, Kien-Thai Yong, Deep Learning for Plant Species Classification Using Leaf Vein Morphometric, IEEE/ACM Trans. Comput. Biol. Bioinform. 17 (1) (2018) 82–90.
[49] Mingle Xu, Hyongsuk Kim, Jucheng Yang, Alvaro Fuentes, Yao Meng, Sook Yoon, Taehyun Kim, Dong Sun Park, Embracing Limited and Imperfect Training Datasets: opportunities and Challenges in Plant Disease Recognition Using Deep Learning, Front Plant. Sci. 14 (2023) 1225409.
[50] B. Yanikoglu, Erchan Aptoula, Caglar Tirkaz, Automatic Plant Identification from Photographs, Mach. Vis. Appl. 25 (2014) 1369–1383.
[51] Ping Zhang, Lihong Xu, Unsupervised Segmentation of Greenhouse Plant Images Based on Statistical Method, Sci. Rep. 8 (1) (2018) 4465.
[52] Xingchen Zhang, Yiannis Demiris, Visible and Infrared Image Fusion Using Deep Learning, IEEE Trans. Pattern. Anal. Mach. Intell. 45 (8) (2023) 10535–10554.
[53] Yuzhuo Zhang, Tianyi Wang, Yong You, Decheng Wang, Jinlong Gao, Tiangang Liang, A Transformer-Based Image Detection Method for Grassland Situation of Alpine Meadows, Comput. Electron. Agric. 210 (2023) 107919.
[54] Li, Shaochun, Yanjun Wang, Hengfan Cai, Yunhao Lin, Mengjie Wang, and Fei Teng. 2023. "MF-SRCDNet: multi-feature fusion super-resolution building change detection framework for multi-sensor high-resolution remote sensing imagery".
[55] M. Tan, Q. Le, Efficientnet: rethinking model scaling for convolutional neural networks, in: International conference on machine learning, PMLR, 2019, May, pp. 6105–6114.