

POL-ID: Deep Learning for Pollen Detection and Classification in the Authentication of South African Honey

Yash Ramklass

rmkyas002@myuct.ac.za

University of Cape Town

Cape Town, Western Cape, South Africa

Abstract

Authenticating the botanical origins of honey is crucial for combating food fraud. Traditional pollen analysis (melissopalynology) is slow and requires specialised expertise, creating a bottleneck for scalable quality control. This research develops and evaluates an end-to-end deep learning pipeline for automating this process in the South African context, guided by three research questions: (i) how accurately modern object detectors, including YOLO variants and DETRs, can localise pollen grains in microscopy slides, (ii) to what extent hybrid CNN–Vision Transformer architectures, implemented through sequential and parallel fusion strategies, can achieve high-accuracy classification of South African pollen grains, and (iii) whether unsupervised clustering of feature embeddings can reliably identify novel or unlabeled pollen types for expert review. Experimental results show that YOLO-based detectors outperform DETRs, with the best model achieving a mean average precision (mAP@50–95) above 0.92, while hybrid classifiers exceed 95% accuracy. The final, optimised pipeline was validated by comparing its analysis of honey-slide images to the findings of expert-led microscopy. This evaluation revealed only a moderate level of alignment: the pipeline produced reasonable honey classifications but showed inconsistencies in identifying dominant taxa. These limitations stem largely from class imbalance in the dataset and error propagation between detection and classification stages. Overall, the system demonstrates strong potential as a scalable tool for honey authentication in South Africa, while also highlighting key areas for improvement.

CCS Concepts

• **Computing methodologies** → **Computer vision; Neural networks**; • **Applied computing** → **Life and medical sciences**.

Keywords

YOLO, DETR, Vision Transformers, CNN, Pollen Classification, Object Detection, Clustering, Deep Learning, Honey Authentication

1 Introduction

The global honey market is susceptible to food fraud, where premium monofloral honeys are often adulterated with cheaper alternatives or their botanical and geographical origins are misrepresented [50]. Authenticating the source of honey is therefore crucial, not only for consumer protection and fair trade but also for preserving the economic integrity of regional apiculture. For South Africa, whose rich and unique biodiversity produces highly sought-after honeys [33], robust authentication methods are essential.

The traditional gold standard for determining honey’s botanical origin is melissopalynology—the microscopic identification

and quantification of pollen grains within a honey sample. This discipline, however, presents a significant bottleneck for scalable honey authentication. The process is slow, labour-intensive, and demands a high level of specialized expertise, which is scarce. Furthermore, the analysis can be subjective, leading to inconsistencies even among trained experts [34].

To overcome these challenges, this research develops and validates an end-to-end deep learning pipeline designed to automate the botanical authentication of South African honey. This work addresses three core research questions. First, how do state-of-the-art object detectors, including YOLO variants and DETRs, perform in localising pollen grains in honey slide images? Second, to what extent can hybrid CNN–Vision Transformer architectures, implemented through sequential and parallel fusion strategies, achieve high-accuracy classification of South African pollen grains? Third, can unsupervised clustering of feature embeddings (using HDBSCAN) reliably identify novel or unlabeled pollen types for expert review? To answer these questions, this study evaluates multiple detection architectures, benchmarks hybrid classification strategies, and integrates clustering into the pipeline. The final system is validated by comparing its palynological profiles with those produced by an expert melissopalynologist. While the models achieve strong performance in isolation, the end-to-end pipeline reveals important limitations, including the effects of heavy class imbalance and error propagation from imperfect detections to subsequent classification. These factors underscore both the promise of automated honey authentication and the need for further refinement.

2 Related Work

The development of automated systems for pollen identification bridges traditional melissopalynology with recent advancements in computer vision and deep learning. This section explores the state of the field, the technical approaches that underpin recent successes, and the specific challenges of South African honey authentication that this work addresses.

2.1 The State of Automated Melissopalynology

Traditional melissopalynology relies on expert visual inspection of microscopic slides, a process prone to error when differentiating between morphologically similar plant species or detecting pollen in filtered honeys [2]. While the International Commission for Bee Botany (ICBB) has introduced methodological standards, inconsistencies between experts remain widespread [9, 34]. Alternative approaches such as chemical profiling and DNA barcoding show promise, but their reliability and cost-effectiveness remain limited [25, 50].

Automated image-based methods using deep learning have demonstrated state-of-the-art performance in pollen recognition across several domains including forensics, allergology, and agriculture [12, 22]. However, existing tools such as Honey.ai and AIPollen are trained on non-African datasets and do not generalize to South African honey samples [49]. Current public datasets remain region-specific, with prominent collections from Spain [30, 47], New Zealand [23], Brazil [13], and India [19]. No large-scale, publicly available dataset of South African honey pollens currently exists, leaving a critical gap in regional honey authentication.

2.2 Technical Approaches in Automated Pollen Analysis

2.2.1 Data Preprocessing and Augmentation. Data quality and scale are consistently identified as key determinants of classification accuracy. Preprocessing techniques such as duplicate removal and the exclusion of blurred images are common [45], while augmentation strategies such as blurring, sharpening, histogram equalization, haze reduction, noise injection, flipping, and rotation are used to increase dataset size and diversity [31, 44, 48].

2.2.2 Detection and Segmentation. Since honey microscopy images typically contain multiple, overlapping grains, an accurate detection stage is essential before classification [39]. A range of approaches have been proposed. Proprietary systems such as MATLAB’s Image Processing Toolbox [35] and the Classifynder system [23, 44] exist, but open-source methods dominate recent work. YOLO-based detectors have been trained for pollen detection and segmentation with high bounding-box accuracy but some limitations in mask generation [14, 20]. CNN-based segmentation architectures such as U-Nets achieve strong performance on complex mixtures but at a high computational cost [3, 48]. Two-stage detectors such as Faster-RCNN and Mask-RCNN have achieved very high detection and segmentation accuracy, particularly when objects are small or overlapping, although they operate more slowly than YOLO [10, 41].

2.2.3 Classification Methods. Pollen classification has advanced from handcrafted feature extraction [46] to deep learning, with CNNs dominating the literature. Both from-scratch models [31, 45] and pre-trained architectures [30, 43] have been tested. While from-scratch CNNs can outperform alternatives when very large datasets are available, pre-trained networks consistently achieve superior results under constrained data conditions [43].

Deep CNNs such as InceptionV3 and ResNeSt-101 have reached accuracies exceeding 97–98% [30, 43], although this comes at significant computational cost. Shallower models like ResNet-18 and VGG16 achieve comparable results in some studies [30, 35], particularly when the number of classes is restricted. Indeed, model accuracy is closely tied to the number of pollen species: performance tends to degrade as class diversity increases, but can be recovered with deeper models or larger datasets [1, 21, 45].

More recently, Vision Transformers (ViTs) have been tested in pollen analysis [17, 19]. ViTs can outperform CNNs on the classification task, though they typically require larger datasets to generalize effectively [37]. Consequently, CNNs remain the architecture of choice for most current pollen classification studies.

2.2.4 Clustering for Novel Pollen Types. Unsupervised clustering offers a means of handling unlabelled or novel pollen species. Daood et al. [7] demonstrate that hierarchical clustering combined with CNNs or LSTMs can improve classification accuracy, while Caron et al. [4] show that alternating clustering with CNN training enables self-supervised learning that rivals supervised baselines. Although clustering alone underperforms CNNs for classification [40], its integration with supervised methods is promising for the authentication of South African honey.

2.3 Identified Gaps and Contributions

The literature reveals several gaps. First, deep learning solutions are currently region-specific, with no publicly available dataset covering the botanically diverse South African context. Second, many existing systems address only classification of pre-cropped grains and neglect the detection stage, limiting their utility for real honey analysis. Third, while CNNs dominate, Transformers remain underexplored in pollen recognition due to their data demands, and clustering techniques have yet to be fully integrated into end-to-end systems.

This project addresses these gaps by: (1) developing the first comprehensive South African pollen dataset in collaboration with the University of Cape Town’s Department of Chemistry; and (2) building and validating a complete pipeline that integrates detection, classification, and clustering, balancing accuracy, adaptability, and computational feasibility.

3 Methodology

This section details the models, software, and hardware utilized in this work. All model training and pipeline scripts were executed on the University of Cape Town (UCT) High-Performance Cluster, using an NVIDIA L40S GPU.

3.1 Datasets and Preprocessing

As this project involved training both a detector and a classifier, it required the creation of two distinct datasets. Both datasets contain light microscopy images of honey slides, which were captured and annotated by collaborators in the Department of Chemistry at UCT.

3.1.1 Detection Dataset. The detection dataset consists of 421 honey slide images, which contain a total of 1330 manually annotated pollen grain bounding boxes. Annotations were created using the Computer Vision Annotation Tool (CVAT). The dataset was partitioned into training (70%), validation (15%), and test (15%) sets. An example of an annotated image is shown in Figure 1.

Prior to training, all images were preprocessed for auto-orientation using Roboflow. Subsequently, to improve generalization and mitigate overfitting, an extensive set of online data augmentations was employed. These can be broadly categorized into three main types:

Geometric augmentations included standard transforms like random horizontal flipping, scaling, and cropping. Photometric augmentations manipulated the color space by randomly adjusting image properties such as hue, saturation, and value to simulate different lighting conditions. Finally, composite augmentations created novel training samples from existing ones. This included Mosaic augmentation, which combines four training images to expose the model to objects at different scales, and MixUp, which generates

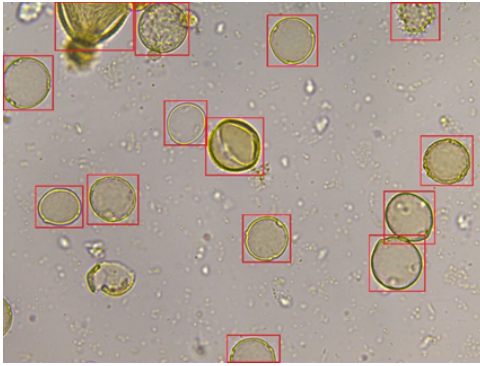


Figure 1: Example of an annotated image from the detection dataset. All pollen grains are enclosed in bounding boxes.

new samples by blending pairs of images and their labels, thereby improving generalization.

3.1.2 Classification Dataset. The classification dataset was constructed from an initial set of 5643 images covering 77 pollen taxa. Reference plates for all taxa can be found in Appendix C. The source images were captured at multiple focal depths, resulting in several images for each unique set of pollen grains. To manage this structure, images corresponding to a single set of grains were grouped into ‘stacks’.

To prevent data leakage, the dataset was partitioned at the stack level into training (70%), validation (15%), and test (15%) sets, ensuring that all images from a given stack belonged exclusively to one split.

Each source image was annotated with a single target taxon, and only grains of that taxon were labeled using CVAT, as shown in Figure 2. A custom script was developed to process these annotations by cropping each labeled pollen grain, thereby creating the final dataset, which consisted of a total of 7594 individual grain images across 710 stacks. While this averaged 9.2 stacks per taxon, the data distribution was highly imbalanced. Stack counts per taxon ranged from a minimum of 3 to a maximum of 29. This imbalance was more pronounced in the final training data: the number of cropped grains per taxon averaged 98.6 with a standard deviation of 83.8, and ranged from a low of 24 to a high of 658.

All cropped pollen images underwent a standardized preprocessing pipeline: they were resized to 224×224 pixels and normalized using the standard ImageNet mean and standard deviation. To enhance model robustness, several online data augmentation strategies were applied during training.

The training augmentation included Random Resized Crop (scaling between 80% and 100%), Random Horizontal and Vertical Flips, Random Rotation (up to 45 degrees), Color Jitter, and Random Erasing.

3.2 Pollen Detection Models

To identify the optimal architecture for pollen grain detection, a comparative study of two major types of modern object detectors was conducted. All evaluated models were pretrained on the COCO dataset [24].

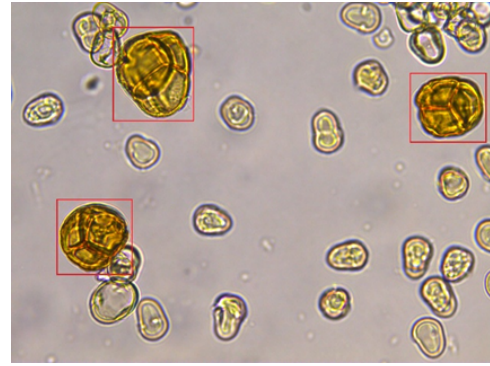


Figure 2: Example of an annotated source image from which individual pollen grains were cropped. Only a single taxon is annotated.

3.2.1 Single-Stage Detectors. Single-stage detectors perform object detection in a single step and are known for their high inference speed [52]. Several prominent variants were evaluated.

The first family of models was YOLO-based, including Ultralytics implementations (YOLOv8 and YOLO11) and YOLOX from the MMDetection library. The Ultralytics models balance speed and accuracy, with the YOLO11s model achieving a mAP@50-95 of 47.0 and inference speeds as low as 2.5ms on COCO [18]. YOLOX was selected for its performance, having offered a better speed-accuracy tradeoff than other state-of-the-art models at its time of release, and its permissive license [11]. An RTMDet model was also trained, which is a real-time detector with improved speed and accuracy over YOLOX and YOLOv7 [29].

The Ultralytics models were trained using their native Python library, while YOLOX and RTMDet were trained using MMDetection. All single-stage detectors were trained with a learning rate of 0.01 and a batch size of 16 for 100–300 epochs, with early stopping to prevent overfitting. The ‘small’ version of each model was initially used to establish a comparable performance baseline, with the nano, small, and medium variants of YOLO11 and YOLOv8 eventually being tested as well.

3.2.2 Transformer-Based Detectors. In contrast to the single-stage approach, two Detection Transformers (DETRs) were also trained. These models reframe object detection as a set prediction problem [52].

The first model was based on the DEIM framework (DETR with Improved Matching), which incorporates a dense matching strategy and novel loss function to boost performance, outperforming models like YOLO11 on COCO [16]. The model variant used was based on D-FINE, a DETR that refines probability distributions for bounding boxes [38], trained using the DEIMKit library [8].

The second model was RF-DETR, a real-time DETR implemented by Roboflow [42]. It is built upon state-of-the-art models including LW-DETR [5], DINOv2 [36], and Deformable DETR [51], achieving high accuracy and speed.

Both DETR models were trained for 200 epochs with a batch size of 16 and a learning rate of 1×10^{-4} . Pre-trained ‘small’ model variants were used in both cases.

3.3 Pollen Classification Models

For pollen classification, two hybrid CNN–Vision Transformer architectures were evaluated. These were implemented using the `timm` library. All models were pretrained on the ImageNet dataset.

3.3.1 Sequential Fusion (CoAtNet). The first architecture was CoAtNet, a sequential fusion model that vertically stacks convolution and attention layers to improve efficiency and generalization [6]. The specific model used was pretrained on the ImageNet-12K dataset and fine-tuned on ImageNet-1K.

3.3.2 Parallel Fusion (ConvNeXt + Swin Transformer). The second architecture was a parallel fusion of a ConvNeXt [27] and a Swin Transformer [26]. ConvNeXt models are pure convolutional networks designed to emulate modern vision transformers, while Swin Transformers use shifted windows to apply self-attention efficiently to local image regions. Both excel at extracting features from local and global contexts.

The parallel fusion model uses ‘small’ Swin and ConvNeXt (both pretrained on ImageNet-1K) backbones to extract feature vectors (size 768 each), which are concatenated into a single 1536-dimensional embedding. This embedding is then passed to a 5-layer MLP classifier head. The head consists of a fully connected layer mapping the 1536-dimensional embedding to 1024 units, followed by batch normalization, a ReLU activation, and a dropout layer with probability 0.5 to reduce overfitting. Finally, it uses another fully connected layer that outputs predictions over the target number of classes. A diagram of the parallel fusion architecture is shown in Figure 3.

3.3.3 Training Strategy. A common two-phase training process was used for both fusion models. This approach represents a standard best practice for transfer learning, built upon principles like gradual unfreezing [15], and is designed to maximize model performance while ensuring training stability. The strategy mitigates the risk of “catastrophic forgetting,” where large gradients from a randomly initialized classifier head can damage the pre-trained weights of the backbone.

- **Phase 1:** The model backbones were frozen, and only the classifier head was trained for 25 epochs with a learning rate of 1×10^{-3} . This initial phase stabilizes the classifier head by allowing it to learn a reasonable mapping from the backbone’s features to the target classes without corrupting the pre-trained weights.
- **Phase 2:** All layers were unfrozen, and the entire model was fine-tuned for 50 epochs. The learning rate for the classifier head was reduced to 5×10^{-4} , while the backbones used a learning rate of 1×10^{-4} . This allows the pretrained backbones to fine-tune their feature representations, aligning them more closely with the statistical distribution of the target pollen data.

A batch size of 16 was used for all experiments. The AdamW optimizer was chosen for its improved generalization in image classification tasks [28]. A cross-entropy loss function was used. This was unweighted in initial experiments, but a weighted version was

also tested to address class imbalance. Further experiments involving automated hyperparameter optimization with RayTune and a three-phase training strategy were also explored.

3.4 Unsupervised Clustering of Unknown Grains

To analyze pollen grains that the classifier could not identify with high confidence, the Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) algorithm [32] was used. HDBSCAN is advantageous as it does not require a predefined number of clusters and can identify clusters of arbitrary shape while effectively managing noise. Clustering was performed on feature embeddings extracted from the classification model, with parameters set to a minimum cluster size of 5 and a minimum samples value of 5, making the algorithm more conservative. Cosine similarity was used to measure distances between feature vectors, as it is effective for high-dimensional data. Pollen grains not assigned to any cluster were labeled as noise, isolating novel groupings from outliers.

3.5 End-to-End Authentication Pipeline

The trained models and clustering algorithm were integrated into a fully automated pipeline to process raw microscope images and generate a detailed palynological profile. The workflow is depicted in Figure 4.

The analysis begins with a set of microscope slide images from a honey sample. The pipeline executes the following sequence:

- (1) **Detection and Cropping:** The detection model scans the images to localize pollen grains. Detections exceeding a confidence threshold D are cropped from the image using their bounding box, with a 10-pixel padding added to ensure the entire grain is captured.
- (2) **Classification:** Each cropped image is passed to the classification model, which extracts a feature embedding and predicts a taxon with a corresponding confidence score.
- (3) **Clustering:** A classification confidence threshold C is applied. Grains with confidence $\geq C$ are considered identified. Grains with confidence $< C$ are labeled ‘unknown’, and their feature embeddings are passed to the unsupervised clustering module.
- (4) **Final Output:** The system aggregates counts from classified and clustered grains to produce a final pollen composition report. This report details the percentage representation of identified taxa and provides a quantitative breakdown of unknown clusters. The final output includes a honey classification, a summary bar chart, a CSV file of the results and exemplar images of clusters for expert review.

The thresholds D and C were calibrated empirically. The optimal detection threshold (D) acts as the primary filter for low-quality detections, helping to discard poor crops. These poor crops typically consist of partial grains located at the image boundary, malformed grains or false positives on non-pollen debris. The classification threshold (C) serves as a secondary filter. Poor crops that pass initial detection often receive low confidence scores from the classifier, so threshold C helps isolate these ambiguous grains for clustering, managing uncertainty while avoiding clustering high-confidence

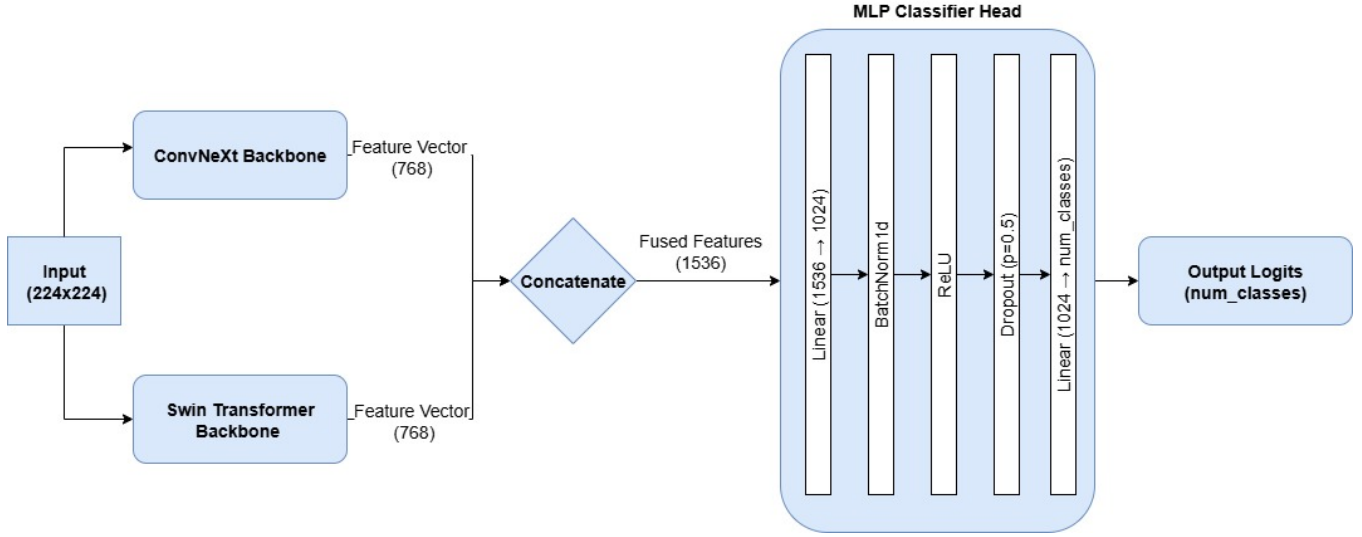


Figure 3: A model architecture diagram illustrating the parallel fusion classification model.

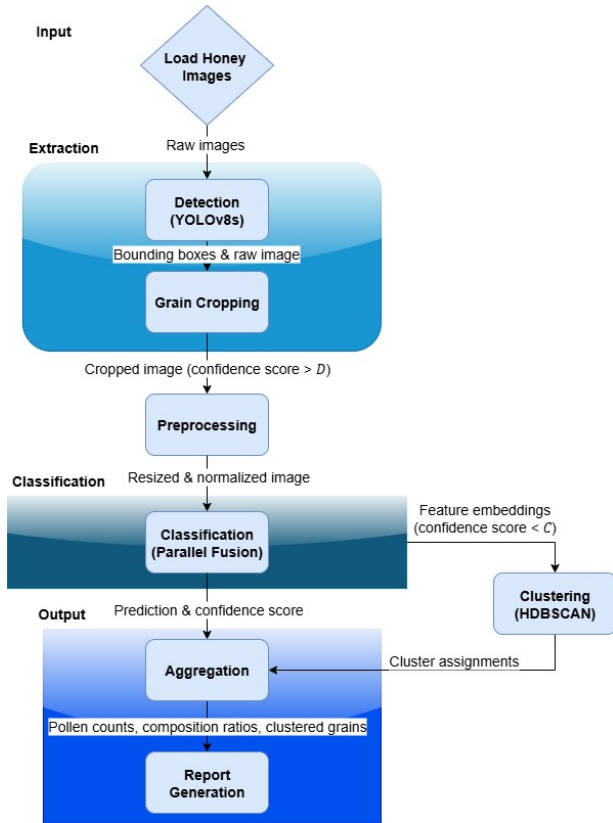


Figure 4: An architecture diagram illustrating the final honey authentication pipeline.

grains. The detailed analysis supporting these choices is presented in Section 4.3.

4 Results and Discussion

This section introduces the metrics used to evaluate the models, clustering algorithm and final pipeline in this study. It presents the evaluation results for each module and discusses their implications for model choice and overall system viability.

4.1 Detection Performance

The detection model for the final pipeline was selected via a two-stage evaluation process. The initial stage involved a comparative analysis of ‘small’ variants from each candidate architecture to identify the most promising model families. In the second stage, a focused analysis on the most successful model families was conducted, evaluating different model sizes to find the best trade-off between accuracy and performance.

4.1.1 Initial Model Evaluation. The initial evaluation of ‘small’ detection models revealed a clear stratification in performance on the pollen grain dataset. As detailed in Table 1, the Ultralytics YOLO models demonstrated superior performance, with both YOLOv8s and YOLO11s achieving mean Average Precision (mAP@50–95) scores above 0.92. Transformer-based architectures performed competently but fell short of this level, with RF-DETR-S reaching 0.894 and DEIM-DFINE-S 0.867. Their underperformance is likely attributable to the limited size of the training dataset (1330 grains), as transformer-based architectures typically require larger amounts of data to converge effectively compared to CNN-based models like YOLO. The MMDetection models, YOLOX-S and RTMDet-S, underperformed substantially with mAP@50–95 scores below 0.78; as relatively older architectures, their weaker performance is not unexpected.

Notably, all models achieved a mAP@50 above 0.92. This indicates that while every architecture could reliably localize pollen grains at the IoU = 0.5 threshold, the Ultralytics models excelled at producing highly precise bounding boxes, as reflected in the stricter mAP@50–95 metric.

Table 1: Comparison of Initial Pollen Detection Models

Model	mAP@50-95	mAP@50	Recall
YOLOv8s	0.9282	0.9875	0.9844
YOLOv11s	0.9201	0.9833	0.9784
RF-DETR-S	0.8944	0.9670	0.9380
DEIM-DFINE-S	0.8670	0.9550	0.9190
YOLOX-S	0.7770	0.9450	0.8360
RTMDet-S	0.7610	0.9230	0.8260

4.1.2 Final Model Selection. Based on the initial results, further evaluation focused on the high-performing Ultralytics models. To determine the optimal balance of accuracy and efficiency, nano, small, and medium variants of YOLOv8 and YOLO11 were benchmarked. The results, summarized in Table 2, highlight a trade-off between model accuracy and computational cost. While YOLOv8s emerged as the most accurate model overall (mAP@50-95 of 0.9282), the YOLO11 family demonstrated remarkable efficiency. For instance, YOLO11n achieved a comparable mAP@50-95 of 0.919 with approximately a quarter of the parameters of YOLOv8s and a 36% faster inference speed.

For the purpose of honey authentication, where real-time inference is not a strict requirement and processing can occur in batches, accuracy was prioritized as the primary selection criterion. Although the accuracy gain of YOLOv8s over YOLO11n is modest (1%), it nonetheless represents the highest performance achieved. Therefore, YOLOv8s was selected as the primary detection model for the end-to-end pipeline. Figure 5 provides a qualitative assessment of the selected model, comparing its predicted bounding boxes against the ground truth labels from the test set. A quantitative view of its classification performance on the single ‘pollen’ class is shown in the normalized confusion matrix in Figure 6. Detailed training graphs for this model can be found in Appendix A.

Looking ahead, the choice of detection model should be guided by deployment constraints. For large-scale applications where throughput or hardware cost become limiting factors, a lighter model such as YOLO11n would offer a compelling balance of accuracy and efficiency. In addition, practical deployment must account for licensing considerations: the permissive Apache-2.0 license of the DETR models could make them preferable to the AGPL-3.0 licensed Ultralytics models in commercial contexts, provided their performance can be enhanced with a larger training dataset or more rigorous tuning.

4.2 Classification Performance

The classification task initially involved 76 taxa across 7 honey samples, with the dataset exhibiting substantial class imbalance. Two architectures were investigated: a sequential fusion model with a CoAtNet backbone pretrained on ImageNet-1k, and a parallel fusion model combining Swin and ConvNeXt backbones pretrained on ImageNet-1k. On this initial dataset, the parallel fusion approach proved considerably stronger, achieving 93.5% accuracy ($F1 = 0.914$) compared to 88.8% accuracy ($F1 = 0.854$) for the sequential model (Table 3). Scaling the parallel fusion from *Tiny* to *Small* yielded further improvements, though moving to *Base* backbones did not provide additional benefit, suggesting diminishing returns from

model size. Additional refinements, including a three-phase training schedule and hyperparameter optimization via RayTune, had minimal effect, indicating that performance was primarily limited by data quality and backbone architecture.

To address these limitations, further pollen images were collected for underrepresented taxa and two taxa were consolidated following expert advice. This expanded the dataset to 77 taxa and reduced imbalance. Experiments on this updated dataset also tested stronger pretraining strategies. The sequential CoAtNet, pretrained on ImageNet-12k and fine-tuned on ImageNet-1k, achieved a marked improvement to 95.8% accuracy ($F1 = 0.939$). In contrast, Swin and ConvNeXt models pretrained on ImageNet-22k did not surpass their 1k-pretrained counterparts in the parallel fusion setting.

A key advance came from addressing class imbalance directly. Replacing the standard Cross Entropy with Weighted Cross Entropy (WCE) had contrasting effects across architectures: the sequential CoAtNet declined in performance (accuracy = 92.9%, $F1 = 0.913$), whereas the parallel fusion improved substantially, reaching 96.5% accuracy and a macro-averaged $F1$ of 0.960 (Table 4). This combination of parallel fusion with WCE represents the best-performing classification model.

Generalization was further assessed by training on subsets of the data. Restricting to three honeys (30 taxa) slightly reduced performance ($F1 = 0.932$), while training on five honeys (54 taxa) improved performance ($F1 = 0.972$). These results suggest that the full 77-taxa model generalizes well, with performance stable across subsets despite residual imbalance.

Finally, efficiency was considered. The sequential CoAtNet achieved competitive results with only 41.7M parameters and 35.5M activations, compared to the larger parallel Swin+ConvNeXt fusion (~100M parameters). Thus, while the parallel model remains the most accurate and is favored in this study, the sequential CoAtNet is a strong candidate for deployment in resource-constrained or latency-sensitive settings.

Training curves and a confusion matrix for the best model (Parallel Swin+ConvNeXt Small-1k with Weighted Cross Entropy) are provided in Appendix A.

4.3 Clustering of Unknown Pollen

Low-confidence grains, defined as those falling below a classifier confidence threshold, were clustered to facilitate expert review. The goal was to group visually similar but uncertain grains, enabling efficient expert verification rather than manual inspection of individual detections. Clustering was performed with HDBSCAN on image embeddings, and evaluated both quantitatively and qualitatively.

Quantitative Evaluation. Clustering quality was assessed using the Silhouette Score (SS) and Davies-Bouldin Index (DBI). Across all honeys, the SS remained between 0.30–0.34, while the DBI ranged from 1.09–1.20, values indicative of stable and reasonably well-separated clusters. A minimum cluster size of 5 was found to prevent fragmented clusters and improve interpretability.

Impact of Detection Threshold. The number of grains available for clustering was influenced by both the detector and classifier

Table 2: Comparison of Ultralytics YOLO Models

Model	mAP@50-95	mAP@50	Recall	Precision	Params (M)	GFLOPS	Speed (ms) Nvidia L40s	Throughput (images/s)
YOLOv8n	0.9240	0.9803	0.9635	0.9674	3.006	8.1	11.2	89.286
YOLOv8s	0.9282	0.9875	0.9844	0.9740	11.126	28.4	14.7	68.027
YOLOv8m	0.9136	0.9701	0.9720	0.9638	25.840	78.7	15.0	66.667
YOLO11n	0.9189	0.9847	0.9740	0.9474	2.582	6.3	10.8	92.593
YOLO11s	0.9201	0.9833	0.9784	0.9495	9.413	21.3	15.2	65.789
YOLO11m	0.9220	0.9858	0.9844	0.9631	20.031	67.6	25.3	39.526

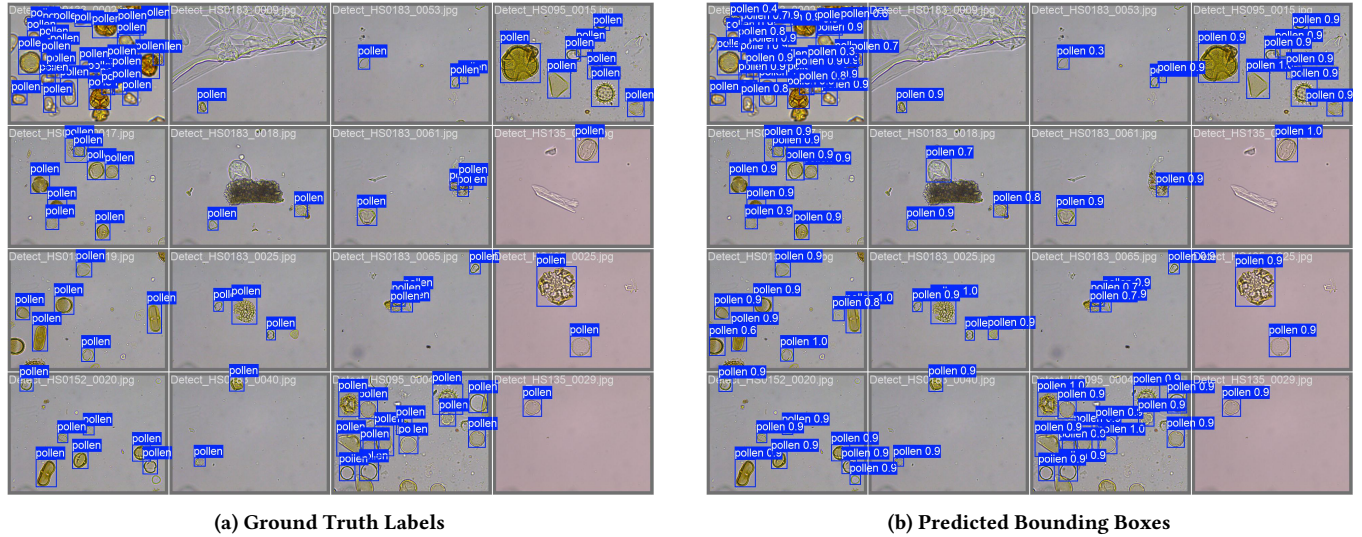


Figure 5: Comparison of ground truth labels and YOLOv8s predicted bounding boxes for a sample from the test set.

Table 3: Initial experiments on the 76-taxa dataset.

Model	Accuracy	Macro Precision	Macro Recall	Macro F1
Sequential CoAtNet (1k)	0.888	0.876	0.867	0.854
Parallel Swin+ConvNeXt (Tiny-1k)	0.935	0.923	0.927	0.914
Parallel Swin+ConvNeXt (Small-1k)	0.933	0.950	0.934	0.923
Parallel Swin+ConvNeXt (Base-1k)	0.933	0.936	0.918	0.910
Parallel (Small-1k, 3-phase)	0.933	0.940	0.922	0.919
Parallel (Small-1k, RayTune)	0.939	0.947	0.930	0.924

Table 4: Final experiments on the 77-taxa dataset.

Model	Accuracy	Macro Precision	Macro Recall	Macro F1
Sequential CoAtNet (12k→1k, CE)	0.958	0.954	0.945	0.939
Parallel Swin+ConvNeXt (Small-1k, CE)	0.932	0.929	0.912	0.906
Sequential CoAtNet (12k→1k, WCE)	0.929	0.931	0.920	0.913
Parallel Swin+ConvNeXt (Small-1k, WCE)	0.965	0.971	0.963	0.960
Parallel Swin+ConvNeXt (Base-1k, WCE)	0.962	0.934	0.949	0.938
Parallel Swin+ConvNeXt (Small-22k→1k, WCE)	0.956	0.963	0.946	0.941

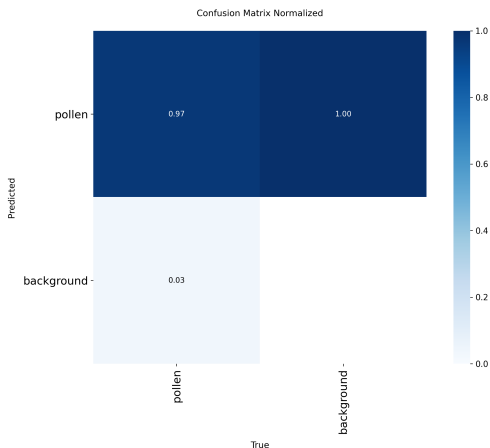


Figure 6: Normalized confusion matrix for the YOLOv8s pollen detector.

thresholds. To calibrate the detector, predicted grain counts were compared against expert counts for multiple thresholds (Table 5). A detection threshold of $D = 0.5$ produced counts most consistent with expert annotations, without introducing large numbers of visually poor crops from low-quality detections.

Table 5: Comparison of expert and detected total grains across detection thresholds.

HoneySample	Expert	$D = 0.5$	$D = 0.6$	$D = 0.7$
HS135	261	283	276	273
HS170	318	365	344	312
HS183	345	278	267	258
HS152	441	349	326	303
HS177	359	341	333	320
HS133	372	304	282	258

Impact of Classifier Threshold. With $D = 0.5$ fixed, the classifier confidence threshold (C) was varied to determine the proportion of grains flagged as low-confidence and clustered. Higher thresholds dramatically increased the number of grains passed to clustering, in some cases exceeding 70% of detections. Figure 7 shows the average proportion of low-confidence grains for each tested threshold. Since the model was generally conservative in assigning confidence, a low threshold ($C = 0.2$) was chosen, resulting in approximately 17% of grains being clustered. This provided a balance between capturing genuinely uncertain grains and avoiding redundant clustering of grains already well-recognized by the model.

Qualitative Evaluation. Visual inspection of outputs (with $D = 0.5$ and $C = 0.2$) further confirmed that clusters captured coherent groups of similar grains. A UMAP projection of embeddings (Figure 8) revealed distinct cluster structure, and exemplar images from

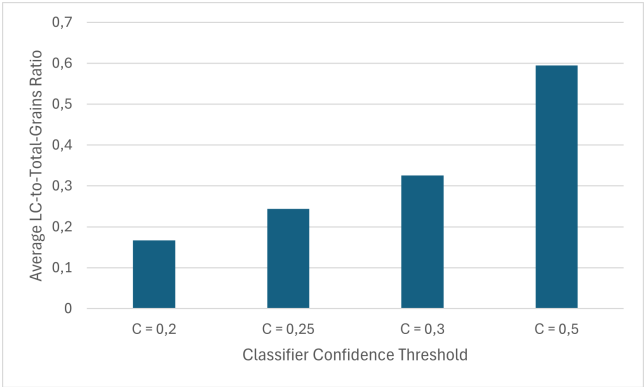


Figure 7: Prevalence of low-confidence grains at different classifier clustering thresholds.

clusters demonstrated consistent morphology within groups (Figure 9). These outputs provide an interpretable pathway for experts to review uncertain grains efficiently.

4.4 Honey Sample Authentication

The complete pipeline combined the best-performing detection model (YOLOv8s), classifier (Parallel Swin+ConvNeXt Small-1k with Weighted Cross Entropy), and clustering thresholds ($D=0.5$, $C=0.2$) as determined earlier. The pipeline was evaluated by comparing model-generated honey classifications to those produced by an expert melissopalynologist. Two sets of expert annotations were used: (i) taxa identified directly from the same pollen grain images supplied to the model, and (ii) conventional melissopalynology using microscopy, where the expert could view slides in three dimensions and adjust focal depth to resolve occluded grains.

Table 6 summarizes monofloral vs. multifloral assignments across methods. A honey was considered monofloral when any taxon exceeded 45% of observed grains; otherwise it was classified as multifloral. Overall, the pipeline's honey classifications show moderate alignment with expert observations. Agreement is strongest in multifloral honeys, where the pipeline consistently matched expert assessments across all six samples. In contrast, the pipeline struggled with monofloral assignments, correctly identifying only one out of three. In borderline cases (e.g., HS095 and HS135), the dominant taxon was underestimated just below the 45% threshold, resulting in multifloral classifications where experts assigned monofloral status. More broadly, the pipeline often fails to reproduce the expert top-3 taxa rankings, highlighting a limitation in capturing finer-grained dominance patterns even when the correct taxa are present in the sample.

To illustrate finer-grained differences, Figure 10 compares pipeline and expert image observations of taxa in a representative honey sample (HS135). While the pipeline correctly identifies *PAL0019* and *PAL0018* as dominant, its estimated relative abundances differ from both expert image and manual observations. Similar charts for all honeys in the study are provided in Appendix B.

These results demonstrate the viability of the pipeline for honey authentication, but also highlight its current limitations. It does not yet reproduce expert top-three rankings with high fidelity, with

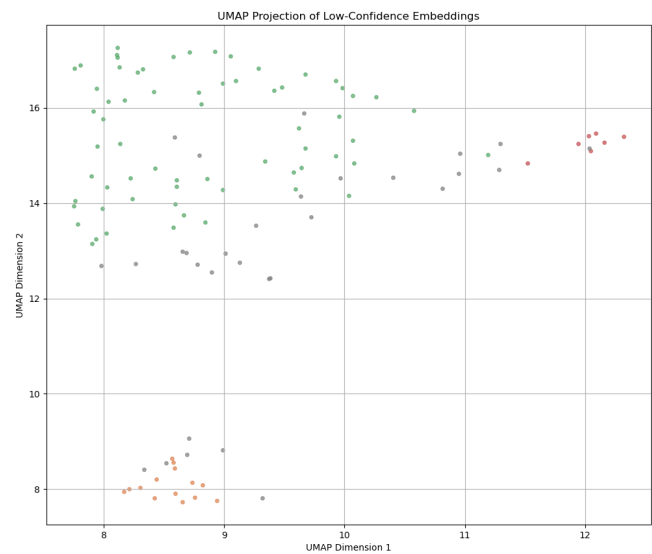


Figure 8: UMAP projection of embeddings colored by HDBSCAN cluster labels (HS170).

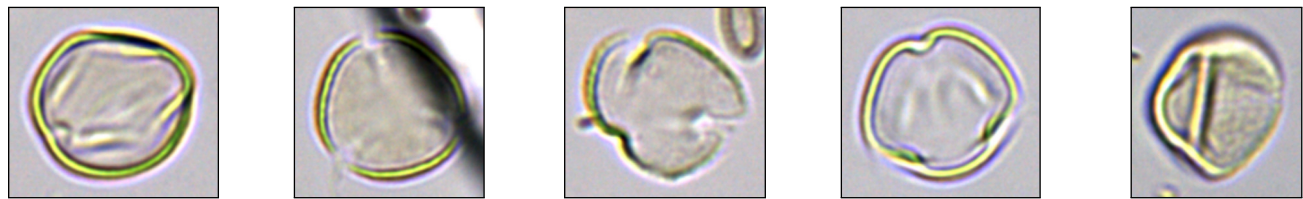


Figure 9: Exemplar images from an HDBSCAN cluster (HS150, cluster 0).

Table 6: Summary of honey classifications across methods. A honey is monofloral when any taxon exceeds 45%; otherwise multifloral. Each cell shows the classification and the dominant (top-abundance) taxon with its proportion.

Honey Sample	Expert Image	Expert Manual	Model
HS095	Monofloral (<i>Celtis</i> , 60.79%)	Monofloral (<i>Celtis</i> , 56.40%)	Multifloral (<i>Celtis</i> , 44.9%)
HS133	Monofloral (<i>Lobostemon sp. 1</i> , 75.8%)	Monofloral (<i>Lobostemon sp. 1</i> , 78.3%)	Monofloral (<i>Lobostemon sp. 1</i> , 56.6%)
HS135	Monofloral (PAL0019, 46.74%)	Monofloral (PAL0019, 56.3%)	Multifloral (PAL0019, 41.0%)
HS152	Multifloral (<i>Apiaceae sp. 1</i> , 26.3%)	Multifloral (<i>Apiaceae sp. 1</i> , 44.0%)	Multifloral (<i>Apiaceae sp. 1</i> , 25.2%)
HS170	Multifloral (<i>Lobostemon sp. 1</i> , 19.8%)	Multifloral (<i>Vahlia sp. 1</i> , 20.0%)	Multifloral (<i>Monocot sp. 2</i> , 13.4%)
HS177	Multifloral (<i>Eucalyptus sp. 3</i> , 26.18%)	Multifloral (<i>Eucalyptus sp. 3</i> , 32.2%)	Multifloral (<i>Eucalyptus sp. 1</i> , 10.3%)
HS183	Multifloral (<i>Eucalyptus sp. 3</i> , 26.08%)	Multifloral (<i>Eucalyptus sp. 3</i> , 27.7%)	Multifloral (<i>Eucalyptus sp. 2</i> , 11.5%)
HS150	Multifloral (PAL0010, 19.36%)	Multifloral (<i>Brassicaceae sp. 2</i> , 19.6%)	Multifloral (<i>Monocot sp. 5</i> , 11.5%)
HS189	Multifloral (PAL0011, 24.72%)	Multifloral (PAL0011, 29.0%)	Multifloral (PAL0011, 25.0%)

divergences most often arising from under- or over-estimation of secondary taxa. Additional discrepancies stem from differences in total grain counts and poor grain crops. These issues indicate a need for further refinement of both the classifier and detection components. Improving these models represents the most promising method of closing the gap to expert-level performance.

5 Conclusion

This study makes several significant contributions to the field of automated palynology, particularly within the South African context. The research successfully established a foundational framework for an automated honey authentication system by developing novel

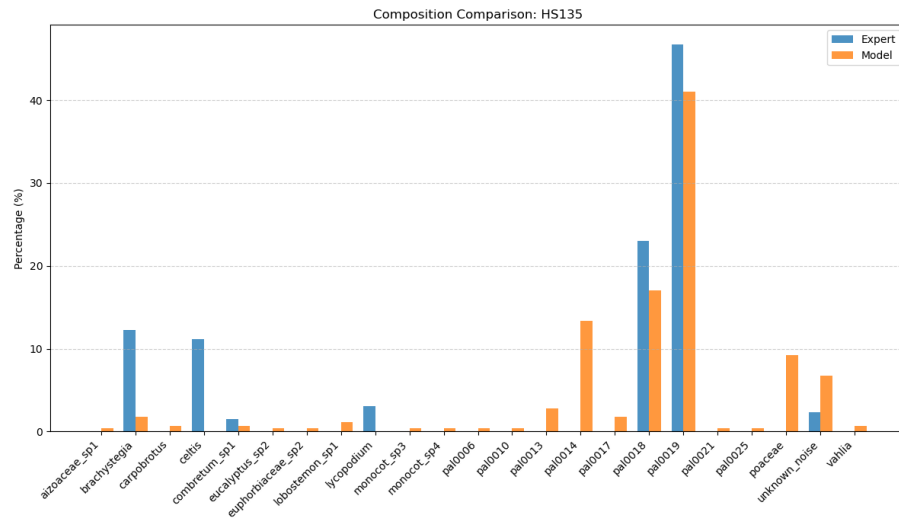


Figure 10: Comparison of expert and model observations on an unseen honey (HS135).

datasets, advanced machine learning models, and an integrated analysis pipeline.

5.1 Key Contributions

The central contribution of this work is the development and validation of a proof-of-concept end-to-end automated honey authentication pipeline tailored to the South African context. This pipeline integrates state-of-the-art machine learning techniques for detection, classification, and clustering of pollen grains, demonstrating the feasibility of automated melissopalynological analysis of South African honeys.

In building this pipeline, several technical advancements were made. A YOLOv8s detection model was trained to localize pollen grains with high precision, achieving a mean Average Precision (mAP@50-95) of 0.9282. For classification, a novel hybrid CNN-Transformer architecture was introduced, achieving 96.5% accuracy across 77 distinct pollen taxa and showing robust scalability to subsets with fewer classes. Additionally, the use of HDBSCAN clustering for unknown or rare grains provided a mechanism for expert review and iterative dataset improvement.

Complementing these modeling efforts, this study developed two novel South African pollen datasets: one for object detection and another for supervised classification. These datasets were critical for training and validating the models and represent valuable resources for future regional research.

Together, these contributions establish the first fully integrated automated honey authentication system in South Africa, providing a proof-of-concept that demonstrates both the feasibility and potential impact of machine learning approaches in palynology.

5.2 Limitations and Future Directions

The principal limitation of this research was the availability of data. Insufficient image samples in both datasets, as well as class imbalance within the final classification dataset constrained model performance. Consequently, the most critical avenue for future

work is the expansion of the datasets, both by increasing the number of annotated images for detection and by adding more images per taxon to improve classification performance and mitigate class imbalance. A focused approach, concentrating on fewer honey types or taxa but with greater data depth, could also yield significant accuracy improvements.

Future work should also enhance the model's generalisation capacity by improving the handling of poor crops. While standard augmentations like random cropping were used, performance could be enhanced by explicitly curating challenging examples, such as partial grains at image boundaries, and using targeted training strategies to force the model to learn useful features from this data.

The scope of model experimentation was also constrained by computational resources and time. Future research would benefit from exploring larger, more complex model architectures and conducting more extensive hyperparameter optimization and data augmentation, contingent on the availability of a larger dataset.

To enhance the practical utility and commercial viability of the pipeline, several key developments are proposed. The implementation of a human-in-the-loop feedback mechanism is a high-priority next step. This system would allow experts to label or correct the classifications of grains identified by the clustering algorithm, facilitating continuous model retraining and improvement.

Furthermore, deploying the pipeline as a web application, similar to platforms such as AIPollen and Honey.AI, would not only facilitate this feedback process but also provide a direct pathway to commercialization. For such a deployment to be successful, optimizing the pipeline for inference speed will be crucial. Future work should therefore investigate the trade-offs between model accuracy and computational speed to develop a system that is both precise and practical for real-world application. Collectively, these future efforts will build upon the foundational work presented herein, advancing the potential for automated palynology to support honey authentication and biodiversity monitoring in South Africa.

Acknowledgments

I would like to express my sincere gratitude to my supervisor, Patrick Marais, for his invaluable guidance and support throughout this project. I also extend my thanks to Cesarina Edmonds-Smith and Janais Delpont from the Department of Chemistry at the University of Cape Town for providing the labeled pollen microscopy images that formed the basis of this work's dataset. Computations were performed using facilities provided by the University of Cape Town's ICTS High Performance Computing team: hpc.uct.ac.za – <https://doi.org/10.5281/zenodo.10021612>.

References

- [1] Gilberto Astolfi, Ariadne Barbosa Gonçalves, Geazy Vilharva Menezes, Felipe Silveira Brito Borges, Angelica Christina Melo Nunes Astolfi, Edson Takashi Matsubara, Marco Alvarez, and Emerson Pistori. 2020. POLLEN73S: An Image Dataset for Pollen Grains Classification. *Ecological Informatics* 60 (November 2020), 101165. doi:10.1016/j.ecoinf.2020.101165
- [2] Nesrin Ecem Bayram. 2020. Molecular approach to detecting pollen types in honey: DNA barcoding. *Hacettepe Journal of Biology and Chemistry* 49, 1 (sep 2020). doi:10.15671/hjbc.623487
- [3] Mihai Boldeanu, Mónica González-Alonso, Horia Cucu, Corneliu Burileanu, Jose Maria Maya-Manzano, and Jeroen Titus Maria Buters. 2022. Automatic Pollen Classification and Segmentation Using U-Nets and Synthetic Data. *IEEE Access* (2022), 73675–73684. doi:10.1109/ACCESS.2022.3189012
- [4] Mathilde Caron, Piotr Bojanowski, Armand Joulin, and Matthijs Douze. 2018. Deep Clustering for Unsupervised Learning of Visual Features. In *Computer Vision – ECCV 2018 (Lecture Notes in Computer Science, Vol. 11218)*, Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss (Eds.). Springer, Cham, 139–156. doi:10.1007/978-3-030-01264-9_9
- [5] Qiang Chen, Xiangbo Su, Xinyu Zhang, Jian Wang, Jiahui Chen, Yunpeng Shen, Chuchu Han, Ziliang Chen, Weixiang Xu, Fanrong Li, Shan Zhang, Kun Yao, Errui Ding, Gang Zhang, and Jingdong Wang. 2024. LW-DETR: A Transformer Replacement to YOLO for Real-Time Detection. arXiv:2406.03459 [cs.CV]
- [6] Zihang Dai, Hanxiao Liu, Quoc V. Le, and Mingxing Tan. 2021. CoAtNet: Marrying Convolution and Attention for All Data Sizes. arXiv:2106.04803 [cs.CV]
- [7] Amar Idrees Daood. 2018. *Pollen Grains Recognition Based on Computer Vision Methods*. Ph. D. Dissertation. Florida Institute of Technology, Melbourne, Florida.
- [8] dnth. 2025. *DEIMKit*. Accessed: August 19, 2025.
- [9] Isabel Escriche, Marisol Juan-Borrás, Mario Visquert, and José Miguel Valiente. 2023. An overview of the challenges when analysing pollen for monofloral honey classification. *Food Control* 143 (jan 2023), 109305. doi:10.1016/j.foodcont.2022.109305
- [10] Ramón Gallardo, Carlos J. García-Orellana, Horacio M. González-Velasco, Antonio García-Manso, Rafael Tormo-Molina, Miguel Macías-Macías, and Eugenio Abengózar. 2024. Automated Multifocus Pollen Detection Using Deep Learning. *Multimedia Tools and Applications* 83, 28 (February 2024), 72097–72112. doi:10.1007/s11042-024-18450-2
- [11] Zheng Ge, Songtao Liu, Feng Wang, Zeming Li, and Jian Sun. 2021. YOLOX: Exceeding YOLO Series in 2021. arXiv:2107.08430 [cs.CV]
- [12] Ariadne Barbosa Gonçalves, Junior Silva Souza, Gercina Gonçalves Da Silva, Marney Pascoli Cereda, Arnildo Pott, Marco Hiroshi Naka, and Emerson Pistori. 2016. Feature Extraction and Machine Learning for the Classification of Brazilian Savannah Pollen Grains. *PLoS ONE* 11, 6 (June 2016), e0157044. doi:10.1371/journal.pone.0157044
- [13] Ariadne Barbosa Gonçalves, Junior Silva Souza, Gercina Gonçalves Da Silva, Marney Pascoli Cereda, Arnildo Pott, Marco Hiroshi Naka, and Emerson Pistori. 2016. Feature Extraction and Machine Learning for the Classification of Brazilian Savannah Pollen Grains. *PLoS ONE* 11, 6 (jun 2016), e0157044. doi:10.1371/journal.pone.0157044
- [14] Chloe He, Alexis Gkantiragas, and Gerard Glowacki. 2018. Honey Authentication with Machine Learning Augmented Bright-Field Microscopy. (2018). arXiv:1901.00516 [cs.CV] doi:10.48550/arXiv.1901.00516
- [15] Jeremy Howard and Sebastian Ruder. 2018. Universal Language Model Fine-tuning for Text Classification. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Melbourne, Australia, 328–339. doi:10.18653/v1/P18-1031
- [16] Shihua Huang, Zhichao Lu, Xiaodong Cun, Yongjun Yu, Xiao Zhou, and Xi Shen. 2024. DEIM: DETR with Improved Matching for Fast Convergence. arXiv:2401.04234 [cs.CV]
- [17] Sonain Jamil, Md. Jalil Piran, and Oh-Jin Kwon. 2023. A Comprehensive Survey of Transformers for Computer Vision. *Drones* 7, 5 (apr 2023), 287. doi:10.3390/drones7050287
- [18] Glenn Jocher and Jing Qiu. 2024. *Ultralytics YOLO11*. <https://github.com/ultralytics/ultralytics>
- [19] Jaidev Sanjay Khalane, Nilesh D. Gawande, Shanmuganathan Raman, and Subramanian Sankaranarayanan. 2025. Advanced Pollen Classification of Indian Medicinal Plants through SEM and Computer Vision. (2025). doi:10.1101/2025.01.08.631879 Preprint.
- [20] Jaidev Sanjay Khalane, Nilesh D. Gawande, Shanmuganathan Raman, and Subramanian Sankaranarayanan. 2025. Advanced Pollen Classification of Indian Medicinal Plants through SEM and Computer Vision. *bioRxiv* (2025). doi:10.1101/2025.01.08.631879
- [21] Natalia Khanzhina, Evgeny Putin, Andrey Filchenkov, and Elena Zamyatina. 2018. Pollen Grain Recognition Using Convolutional Neural Network. In *Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2018)*. i6doc.com, Bruges, Belgium. <http://www.i6doc.com/en/>
- [22] Ewa Kubera, Agnieszka Kubik-Komar, Krystyna Piotrowska-Weryszko, and Małgorzata Skrzypiec. 2021. Deep Learning Methods for Improving Pollen Monitoring. *Sensors* 21, 10 (May 2021), 3526. doi:10.3390/s21103526
- [23] Ryan Lagerstrom, Katherine Holt, Yulia Arzhaeva, Leanne Bischof, Simon Haberle, Felicitas Hopf, and David Lovell. 2015. Pollen Image Classification Using the Classifyfyder System: Algorithm Comparison and a Case Study on New Zealand Honey. In *Signal and Image Analysis for Biomedical and Life Sciences*, Changming Sun, Tomasz Bednarz, Tuan D. Pham, Pascal Vallotton, and Dadong Wang (Eds.). Springer International Publishing, Cham, 207–226. doi:10.1007/978-3-319-10984-8_12
- [24] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. 2014. Microsoft COCO: Common objects in context. In *European conference on computer vision*. Springer, 740–755.
- [25] Shanlin Liu, Dandan Lang, Guanliang Meng, Jiahui Hu, Min Tang, and Xin Zhou. 2022. Tracing the origin of honey products based on metagenomics and machine learning. *Food Chemistry* 371 (mar 2022), 131066. doi:10.1016/j.foodchem.2021.131066
- [26] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. 2021. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. arXiv:2103.14030 [cs.CV]
- [27] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. 2022. A ConvNet for the 2020s. arXiv:2201.03545 [cs.CV]
- [28] Ilya Loshchilov and Frank Hutter. 2017. Decoupled Weight Decay Regularization. arXiv:1711.05101 [cs.LG]
- [29] Chengqi Lyu, Wenwei Zhang, Haian Huang, Yue Zhou, Yudong Wang, Yanyi Liu, Shilong Zhang, and Kai Chen. 2022. RTMDet: An Empirical Study of Designing Real-Time Object Detectors. arXiv:2212.07784 [cs.CV]
- [30] Fernando López-García, José Miguel Valiente, Isabel Escriche Roberto, Maria del Surfrago Juan-Borrás, Marisol Visquert Fas, Verónica Atienza-Vanacloig, and Marta Agustí-Melchor. 2023. Classification of Honey Pollens with ImageNet Neural Networks. In *Lecture Notes in Computer Science*, Vol. 14185. 192–200. doi:10.1007/978-3-031-44240-7_19
- [31] Tahir Mahmood, Jiho Choi, and Kang Ryoung Park. 2023. Artificial Intelligence-Based Classification of Pollen Grains Using Attention-Guided Pollen Features Aggregation Network. *Journal of King Saud University - Computer and Information Sciences* 35, 2 (February 2023), 740–756. doi:10.1016/j.jksuci.2023.01.013
- [32] Leland McInnes, John Healy, and Steve Astels. 2017. hdbscan: Hierarchical density based clustering. *Journal of Open Source Software* 2, 11 (2017), 205. doi:10.21105/joss.00205
- [33] Palesa Mothapo, Sakhile Tsotsobe, Terence P. Jayiya, and Tlou Masehela. 2018. BUSINESS CASE FOR WILD BEE POPULATIONS IN THE WESTERN CAPE: BASELINE ASSESSMENT - Value, Ecological Risks and Amelioration. Technical Report. South African National Biodiversity Institute. doi:10.13140/RG.2.2.21101.03044
- [34] Werner Von Der Ohe, Livia Persano Oddo, Maria Lucia Piana, Monique Morlot, and Peter Martin. 2004. Harmonized methods of melissopalynology. *Apidologie* 35, Suppl. 1 (2004), S18–S25. doi:10.1051/apido:2004050
- [35] Ola Olsson, Melanie Karlsson, Anna S. Persson, Henrik G. Smith, Vidula Varadarajan, Johanna Yourstone, and Martin Stjernman. 2021. Efficient, Automated and Robust Pollen Analysis Using Deep Learning. *Methods in Ecology and Evolution* 12, 5 (May 2021), 850–862. doi:10.1111/2041-210X.13575
- [36] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, Mahmoud Assran, Nicolas Ballas, Wojciech Galuba, Russell Howes, Po-Yao Huang, Shang-Wen Li, Ishan Misra, Michael Rabbat, Vasu Sharma, Gabriel Synnaeve, Hu Xu, Hervé Jegou, Julien Mairal, Patrick Labatut, Armand Joulin, and Piotr Bojanowski. 2023. DINOv2: Learning Robust Visual Features without Supervision. arXiv:2304.07193 [cs.CV]
- [37] Sayak Paul and Pin-Yu Chen. 2022. Vision Transformers Are Robust Learners. *AAAI* 36, 2 (June 2022), 2071–2081. doi:10.1609/aaai.v36i2.20103
- [38] Yansong Peng, Hebei Li, Peixi Wu, Yueyi Zhang, Xiaoyan Sun, and Feng Wu. 2024. D-FINE: Redefine Regression Task in DETRs as Fine-grained Distribution Refinement. arXiv:2403.13842 [cs.CV]

- [39] Surangi W. Punyasena, Derek S. Haselhorst, Shu Kong, Charless C. Fowlkes, and J. Enrique Moreno. 2022. Automated Identification of Diverse Neotropical Pollen Samples Using Convolutional Neural Networks. *Methods in Ecology and Evolution* 13, 9 (September 2022), 2049–2064. doi:10.1111/2041-210X.13917
- [40] A. Rebiai, B. Ben Seghir, H. Hemmami, S. Zeghoud, T. Siham, I. Kouadri, H. Terea, and F. Brahmia. 2021. Clustering and discernment of Algerian bee pollen using an image analysis system. *Algerian Journal of Chemical Engineering* 2, 2 (2021), 41–48. doi:10.5281/zenodo.5105756
- [41] Luca Rettenberger, Friedrich Rieken Münke, Roman Bruch, and Markus Reischl. 2023. Mask R-CNN Outperforms U-Net in Instance Segmentation for Overlapping Cells. *Current Directions in Biomedical Engineering* 9, 1 (September 2023), 335–338. doi:10.1515/cdbme-2023-1084
- [42] Isaac Robinson, Peter Robicieux, and Matvei Popov. 2024. *RF-DETR*. SOTA Real-Time Object Detection Model.
- [43] Masoud A. Rostami, Behnaz Balmaki, Lee A. Dyer, Julie M. Allen, Mohamed F. Sallam, and Fabrizio Frontalini. 2023. Efficient Pollen Grain Classification Using Pre-Trained Convolutional Neural Networks: A Comprehensive Study. *Journal of Big Data* 10, 1 (October 2023), 151. doi:10.1186/s40537-023-00815-3
- [44] Victor Sevillano, Katherine Holt, and José L. Aznarte. 2020. Precise Automatic Classification of 46 Different Pollen Types with Convolutional Neural Networks. *PLoS ONE* 15, 6 (June 2020), e0229751. doi:10.1371/journal.pone.0229751
- [45] F. M. Javed Mehedi Shamrat, Mohd Yamani Idna Idris, Xujuan Zhou, Majdi Khalid, Sharmin Sharmin, Zeseya Sharmin, Kawsar Ahmed, and Mohammad Ali Moni. 2024. PollenNet: A Novel Architecture for High Precision Pollen Grain Classification through Deep Learning and Explainable AI. *Heliyon* 10, 19 (October 2024). doi:10.1016/j.heliyon.2024.e38596
- [46] Diogo Soares, Diego Silva, Lia Quinta, Hemerson Pistori, and Marcelo Borth. 2014. Application of wavelet transform in the classification of pollen grains. *African Journal of Agricultural Research* 9 (2014), 908–913. doi:10.5897/AJAR2013.749
- [47] José Miguel Valiente, Marisol Juan-Borrás, Fernando López-García, and Isabel Escriche. 2023. Automatic Pollen Recognition Using Convolutional Neural Networks: The Case of the Main Pollens Present in Spanish Citrus and Rosemary Honey. *Journal of Food Composition and Analysis* 123 (October 2023), 105605. doi:10.1016/j.jfca.2023.105605
- [48] Nana Yang, Victor Joos, A.-L. Jacquemart, Christel Buyens, and C. De Vleeschouwer. 2022. Using Pure Pollen Species When Training a CNN to Segment Pollen Mixtures. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 1694–1703. doi:10.1109/CVPRW56347.2022.00176
- [49] X. Yu, J. Zhao, Z. Xu, J. Wei, Q. Wang, F. Shen, X. Yang, and Z. Guo. 2024. AIPollen: An Analytic Website for Pollen Identification Through Convolutional Neural Networks. *Plants* 13, 22 (2024), 3118. doi:10.3390/plants13223118
- [50] Xiao-Hua Zhang, Hui-Wen Gu, Ren-Jun Liu, Xiang-Dong Qing, and Jin-Fang Nie. 2023. A comprehensive review of the current trends and recent advancements on the authenticity of honey. *Food Chemistry: X* 19 (oct 2023), 100850. doi:10.1016/j.fochx.2023.100850
- [51] Xizhou Zhu, Weijie Su, Lewei Lu, Bin Li, Xiaogang Wang, and Jifeng Dai. 2020. Deformable DETR: Deformable Transformers for End-to-End Object Detection. arXiv:2010.04159 [cs.CV]
- [52] Zhengxia Zou, Zhenwei Shi, Yuhong Guo, and Jieping Ye. 2019. Object Detection in 20 Years: A Survey. *CoRR* abs/1905.05055 (2019). arXiv:1905.05055 <http://arxiv.org/abs/1905.05055>

Appendices

Appendix A Detailed Training Metrics

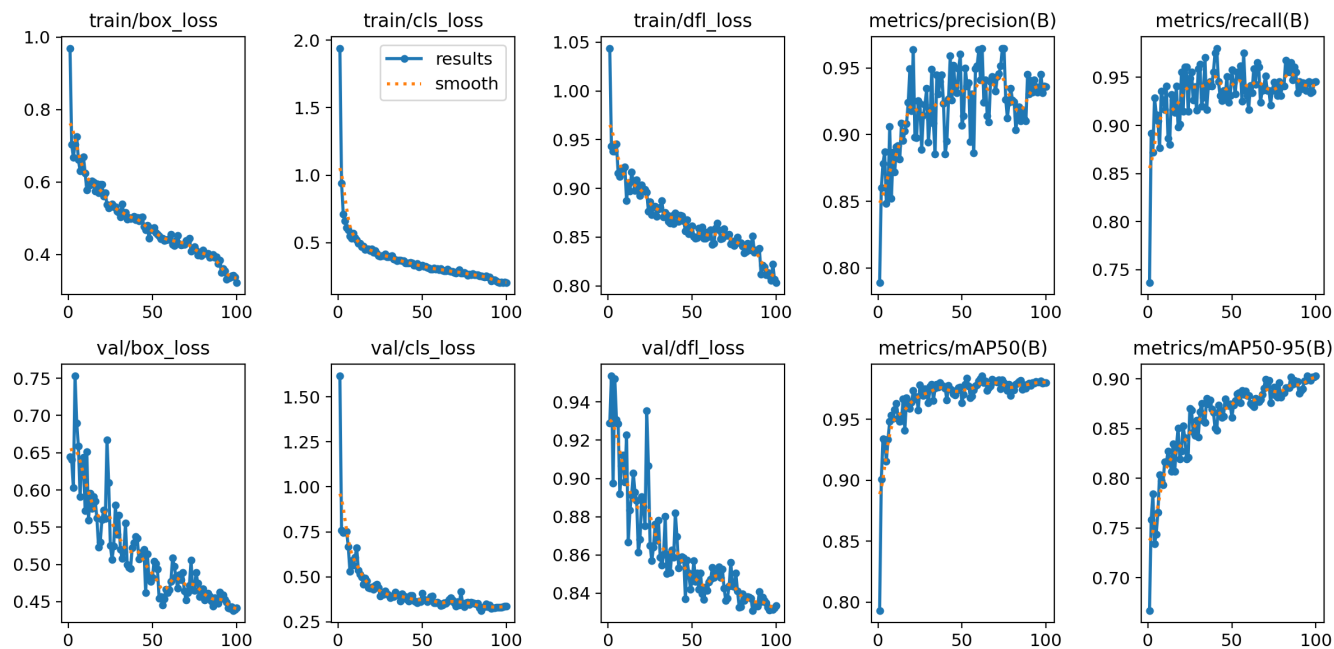
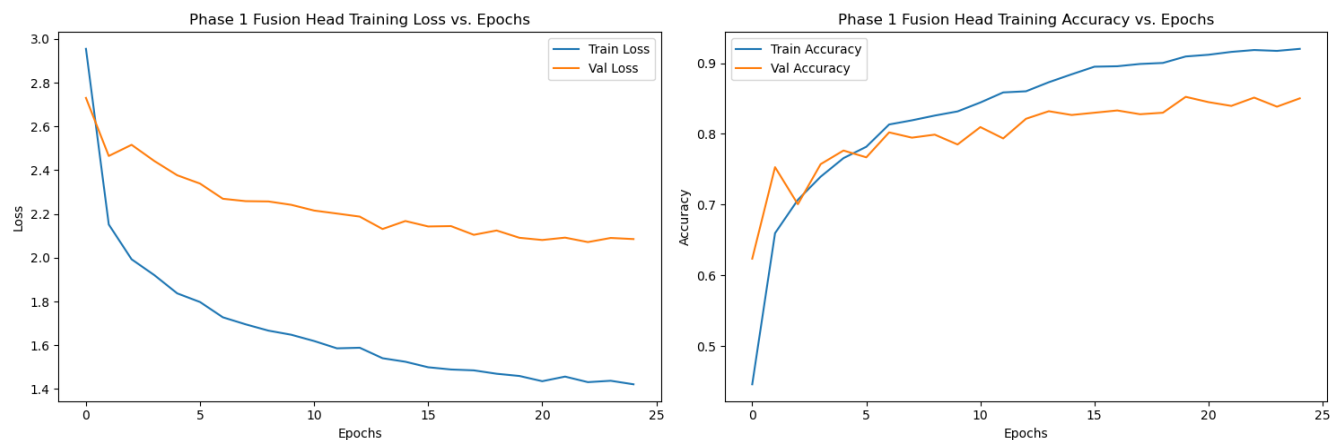
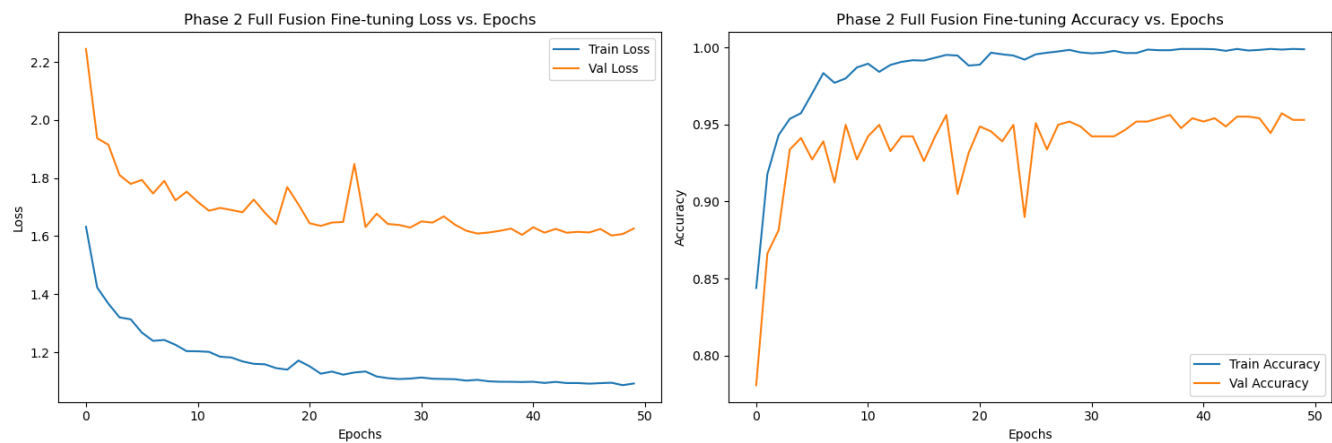


Figure 11: Training and validation performance graphs for the YOLOv8s detection model.



(a) Phase 1 training graphs



(b) Phase 2 training graphs

Figure 12: Phase 1 and 2 training graphs for the final parallel fusion classification model.

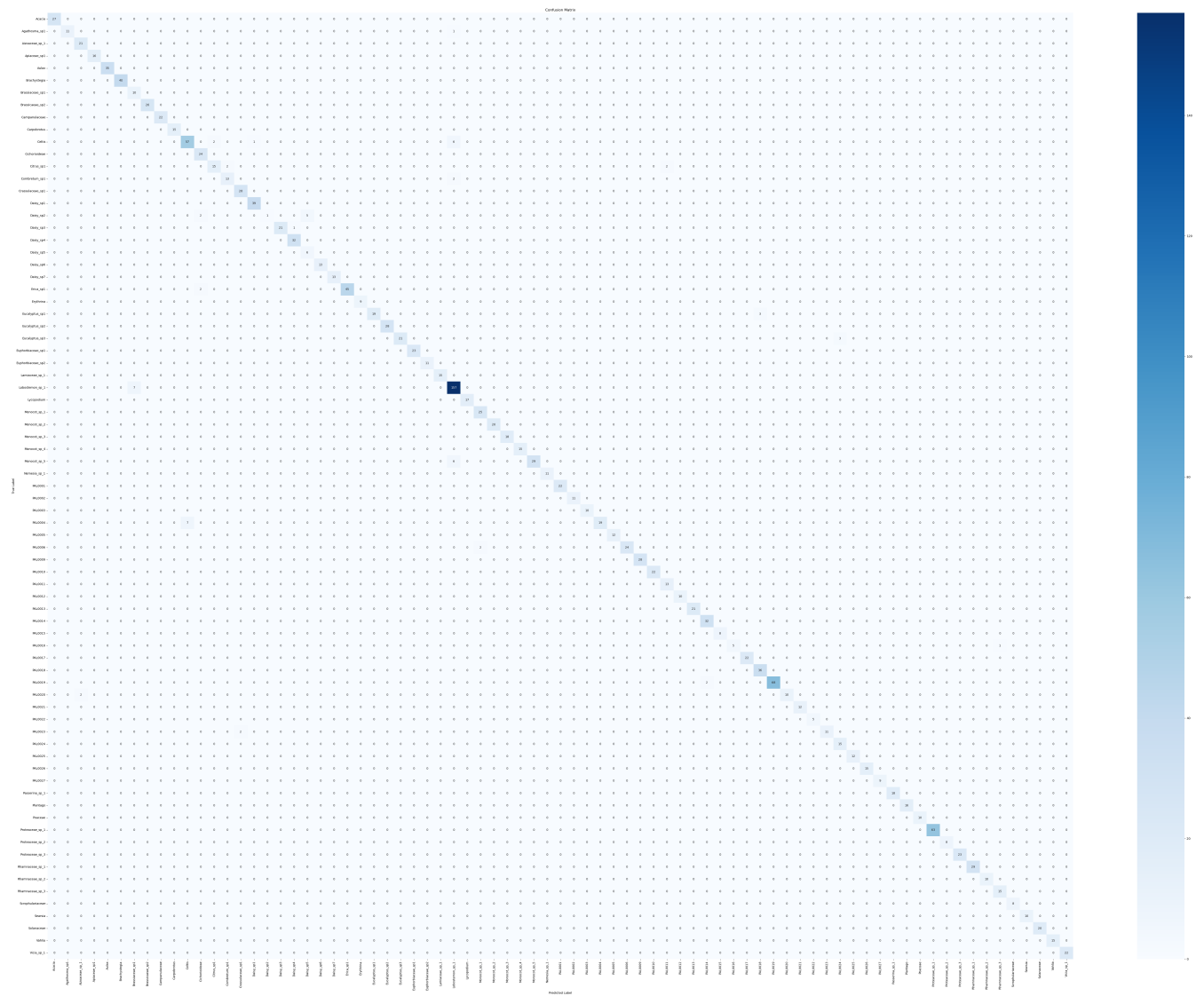


Figure 13: Confusion matrix for the final parallel fusion classification model.

Appendix B Detailed Comparison of Results

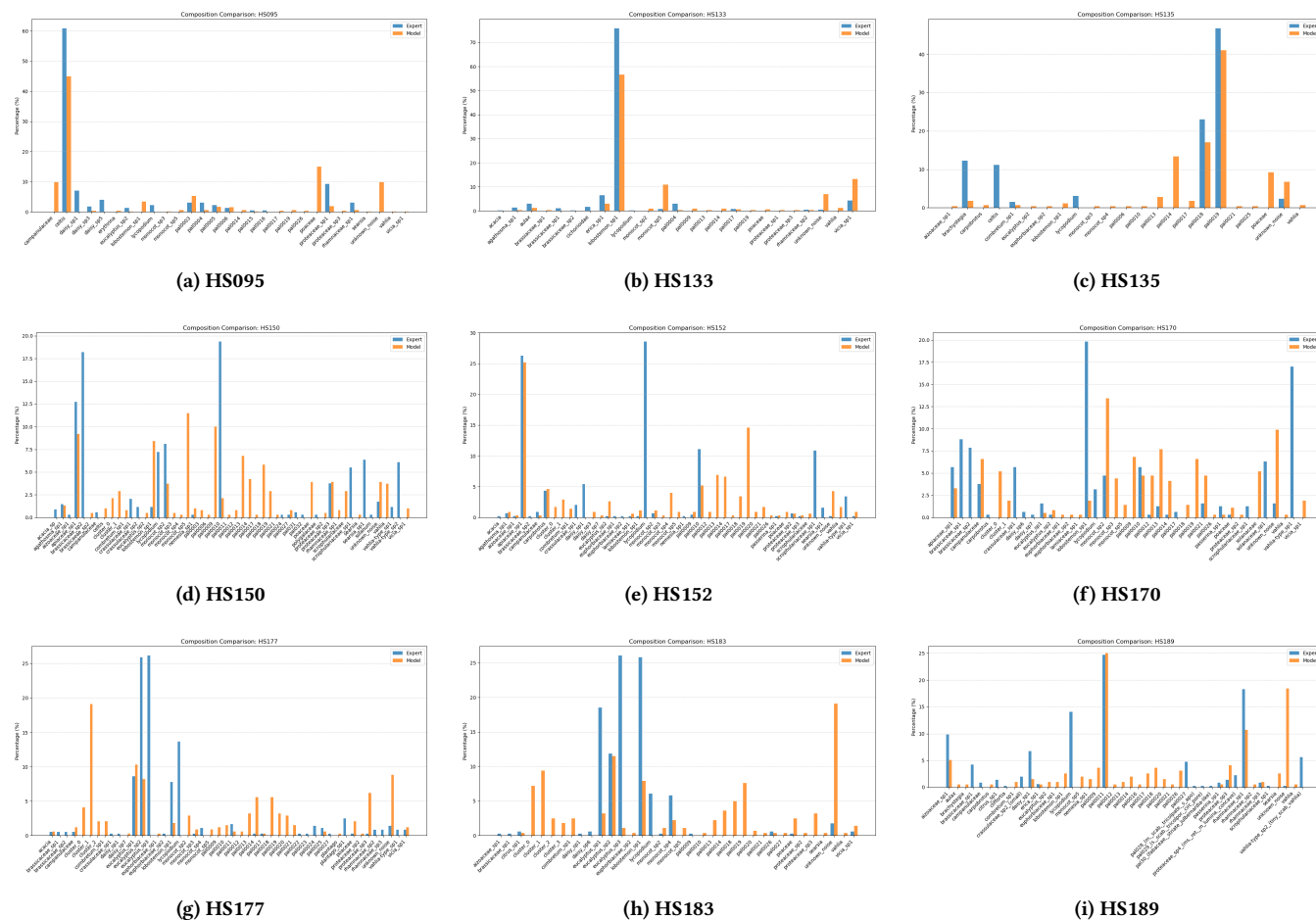


Figure 14: Comparison of expert and model observations on nine unseen honey samples.

Appendix C Pollen Reference Plates

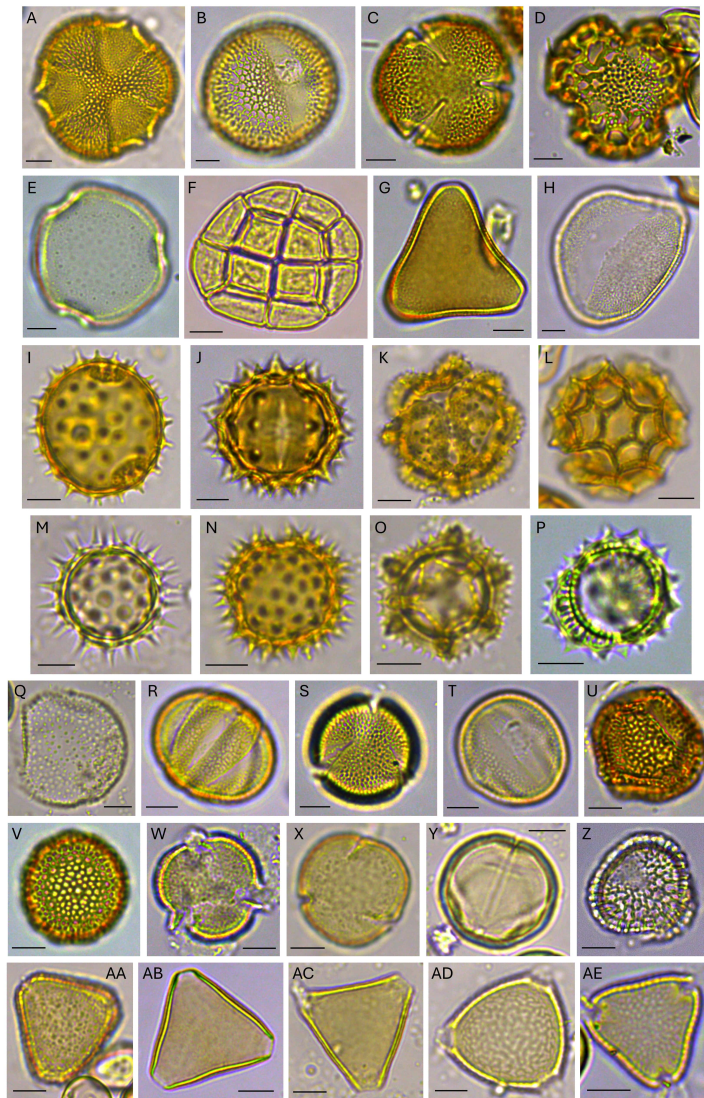


Figure 15: A PAL0001, B PAL0012, C Euphorbiaceae sp. 1, D Brachystegia, E Campanulaceae, F Acacia sp. 1, G Proteaceae sp. 2, H Monocot sp. 3, I PAL0002, J Daisy sp. 7, K Daisy sp. 2, L Cichorioideae, M Daisy sp. 3, N Daisy sp. 4, O Daisy sp. 5, P Daisy sp. 6, Q Monocot sp. 1, R Lamiaceae sp. 1, S PAL0009, T PAL0021, U PAL0015, V Thymelaeaceae, W Euphorbiaceae sp. 2, X PAL0006, Y PAL0003, Z PAL0022, AA Aulax, AB Proteaceae sp. 3, AC Proteaceae sp. 1, AD Erythrina, AE PAL0026.



Figure 16: : A PAL0014, B PAL0019, C Solanaceae sp. 1, D PAL0023, E Erica sp. 1, F Poaceae, G PAL0013, H Rhamnaceae sp. 1, I Rhamnaceae sp. 2, J Rhamnaceae sp. 3, K PAL0020, L Eucalyptus sp. 1, M Eucalyptus sp. 2, N Lycopodium, O Carpobrotus, P PAL0017, Q Plantago sp. 1, R Searsia sp. 1, S PAL0018, T PAL0027, U Celtis, V PAL0005, W Crassulaceae sp. 1, X Aizoaceae sp. 1, Y Citrus sp. 1, Z Brassicaceae sp. 2, AA Brassicaceae sp. 1, AB PAL0011, AC Daisy sp. 1, AD PAL0025, AE PAL0016, AF Apiaceae sp. 1, AG Agathosma sp. 1, AH Vicia sp. 1, AI Monocot sp. 2, AJ Vahlia-type sp. 1, AK Scrophulariaceae sp. 1, AL PAL0024, AM Monocot sp. 5, AN PAL0004, AO Monocot sp. 4, AP PAL0010, AQ Eucalyptus sp. 3, AR Lobostemon sp. 1, AS Nemesia sp. 1, AT Combretum sp. 1.