POL-ID: Automated Honey Authentication Through Deep Learning-Based Pollen Grain Analysis

Maryam Mather
University of Cape Town
Rondebosch, Western Cape, South Africa
mthmar046@myuct.ac.za

ABSTRACT

Honey authentication is critical for food safety and preventing economic fraud in the global honey market. Traditional melissopalynological analysis remains the standard for determining honey origin and authenticity but is time-consuming and requires specialized expertise. This study presents POL-ID, an automated honey authentication pipeline that combines deep learning-based pollen grain classification with established melissopalynological standards from the International Commission for Bee Botany (ICBB).

The system integrates four key components: YOLOv11-nano for pollen grain detection (achieving 97.73% mAP@0.5), ConvNeXt Tiny for species classification across 77 pollen taxa (93.51% accuracy), HDBSCAN clustering for novel pollen type discovery (ARI = 0.962), and an automated honey classification pipeline. The approach addresses challenges in pollen analysis including class imbalance, morphological similarity, and uncertainty quantification through confidence-based routing and unsupervised clustering.

Testing on South African honey samples, POL-ID successfully classified honey types with high accuracy despite underestimation of dominant pollen percentages. The system processes complete honey slides automatically, providing detailed pollen distribution analysis, confidence scoring, and ICBB-compliant honey authentication reports. POL-ID demonstrates the potential for AI-driven automation in honey authentication, which offers a scalable solution for quality control while reducing analysis time and expertise requirements. The system provides a foundation for applications in botanical origin verification in South African apiculture.

CCS CONCEPTS

• **Computing methodologies** → **Machine learning**; **Neural networks**; **Computer vision**.

KEYWORDS

honey authentication, pollen classification, deep learning, computer vision, food fraud detection, melissopalynology, YOLO, ConvNeXt, clustering, HDBSCAN

1 INTRODUCTION

Pollen analysis is central to honey quality control, as it verifies both botanical origin and product authenticity [1]. Traditionally, this analysis is done through melissopalynology, a manual process that is labour-intensive, time-consuming and often costly [14]. It can also be prone to human error and have subjective interpretations especially the differentiation of pollen grains with very subtle morphological differences [10]. The South African honey industry highlights these challenges as the biodiversity of the regions produces a wide variety of pollen taxa, making authentication vital but complex.

To overcome these limitations, there has been a growing demand for the automation of pollen analysis. Artificial Intelligence (AI) and machine learning (ML), particularly deep learning (DL), have emerged as promising solutions to automate the pollen classification directly from microscopic images [22]. Deep learning models, like Convolutional Neural Networks (CNNs), are well-suited for image-based classification as they can automatically determine and extract discriminative features from images [27]. This will allow for enhanced efficiency and speed, improved accuracy and non-reliance on specialized expertise. Whilst these methods have been applied in other countries with their region-specific taxa, there remains a gap of honey authentication tools designed to specifically handle South African pollen diversity.

This paper aims to build the first automated honey authentication system designed for South African pollen. The goal was to achieve classification accuracy of 80-85% or higher to provide reliable benchmarks for honey authentication in South Africa. To reach this, we focused on four areas. First, we aimed to develop and validate a CNN-based classification approach that can handle the exceptional taxonomic diversity of South African pollen and combine it with unsupervised clustering to flag unknown or novel types. A YOLO-based detection module was added to locate pollen grains on slides. Finally, we aimed to identify and implement optimal techniques for handling class imbalance and morphological similarity in diverse pollen datasets.

This study makes a number of contributions. It introduces the first comprehensive automated honey authentication system tailored to South African pollen taxa. It demonstrates a CNN-based approach that achieves strong classification accuracy, with the integration of unsupervised clustering for the discovery of novel pollen types. Finally, it presents an automated pipeline for honey classification, establishing a foundation for both scientific research and industry application to verify if the pollen complies with the declared botanical origin of a honey sample [42].

2 BACKGROUND AND RELATED WORK

2.1 Background

2.1.1 Melissopalynology and Honey Authentication. Melissopalynology is the manual study of pollen grains in honey to determine its botanical and geographical origin [41]. It is a traditional method of honey analysis which requires skilled experts which can be both labour intensive and time consuming [27]. The method has long relied on microscopic identification and quantitative analysis of pollen, following the standards of the ICBB [35].Within this system,

honey is described as monofloral when at least 45% of the grains belong to one plant type, as mixed when the dominant type falls between 20 and 44%, and as multifloral when no single type exceeds 20% [32].

Preparing and analysing a single sample often takes several hours [9]. South Africa's flora adds a further challenge. With more than 24,000 plant species, the pollen diversity far exceeds that of most honey-producing regions, which makes the task of authentication especially demanding [24].

2.1.2 Computer Vision Fundamentals. Computer vision has offered new ways to approach pollen analysis. Object detection methods such as YOLO (You Only Look Once) predict both the location and identity of objects in a single step [13]. It divides an image into grids and regresses bounding boxes directly. Modern versions, such as YOLOv8 and beyond, introduce stronger feature pyramids and anchor-free detection heads to improve accuracy [39].

Convolutional neural networks (CNNs) form the backbone of most vision tasks. They learn layered feature representations through successive convolution and pooling operations [38]. Modern designs such as ConvNeXt combine established convolutional structures with ideas from vision transformers, using depthwise separable convolutions and layer normalization to balance accuracy with efficiency [18].

Transfer learning makes these networks practical for domains where data is limited. By starting from models pre-trained on large datasets such as ImageNet, the lower layers can supply general visual features while higher layers are tuned to the specific classification problem [40]. Performance is then measured with standard metrics. Precision quantifies the proportion of correct positive predictions [1], while recall captures how many actual positives are detected [22]. The F1-score balances the two as a harmonic mean of precision and recall [19]. For detection, mean Average Precision (mAP) at different overlap thresholds provides the benchmark, with mAP@0.5 being the most common measure. [15].

2.1.3 Unsupervised Learning and Clustering. Clustering methods group similar pollen samples without labels, helping to identify patterns or novel types [1]. HDBSCAN, a hierarchical extension of DBSCAN, builds clusters of varying density and identifies points that do not fit any group. This makes it suitable for datasets where the number of clusters is unknown in advance [2].

Before clustering, high-dimensional features from deep networks often require dimensionality reduction. Principal Component Analysis (PCA) offers a linear method that retains as much variance as possible in fewer dimensions. UMAP provides a non-linear alternative that preserves local structure, making it effective for both visualization and preprocessing [25].

To judge the quality of clustering, several metrics are used. The Adjusted Rand Index (ARI) compares discovered clusters with ground-truth labels, ranging from -1 to 1, with higher values indicating stronger agreement [36]. The silhouette coefficient measures how distinct the clusters are relative to one another [29], while the Davies–Bouldin index evaluates the ratio of within-cluster similarity to between-cluster separation, with lower scores reflecting better quality [6].

2.1.4 Class Imbalance and Deep Learning Optimization. A recurring challenge in pollen datasets is class imbalance. Some taxa are abundant, while others appear only rarely. Techniques such as focal loss reduce the influence of easy examples and focus the model on minority classes [17]. Weighted sampling increases the chance of underrepresented classes being selected during training. Stratified splitting maintains balanced proportions across training, validation, and test sets.

Data augmentation further strengthens training by simulating the natural variability of pollen. Geometric changes such as rotations, flips, and crops help the model cope with differences in orientation and position [30]. Photometric adjustments, like brightness and contrast, account for variation in imaging conditions. These transformations must be chosen carefully so as to not erase the visual traits that distinguish one pollen type from another.

2.2 Related Work

2.2.1 Manual methods and in the South African Landscape. In addition to melissopalynology, conventional analyses such as physicochemical and sensory methods have been widely used to ensure honey authenticity [42]. However, these methods have gradually been abandoned due to results variation and the continuous development of adulteration strategies. In South Africa, recent works that regard melissopalynology follow conventional light-microscope pollen analysis to trace honey's botanical origin [26]. It also involves analyzing parameters such as sugars (fructose, glucose), pH, total acidity, moisture, and ash to monitor honey quality. However, South Africa's unique flora makes pollen identification difficult, and incidental wind-blown pollen (e.g. grasses) can skew honey spectra. There have been no works in a South African context that relate to an automated pollen identification system, which highlights the need for pollen monitoring systems to overcome the limitations of manual methods, with such systems being developed and tested across various regions [1] [5].

2.2.2 Automated Pollen Detection. Recent studies have introduced the concept of automated pollen detection which aim to overcome the limitations of manual analysis, making the process faster. A novel approach for identifying fraudulent honey has been developed that utilizes machine learning augmented microscopy [11]. This system was specifically designed to segment and identify the botanical origin, and distribution of pollen grains from microscopes. A three-class YOLOv2 network was trained to detect and segment pollen. It demonstrated promising results as performance metrics included a precision of 0.663, sensitivity/recall of 0.914, and an F1-score of 0.769. Air bubbles formed the majority of false positives.

2.2.3 Automated Honey Classification. Deep learning, particularly CNNs, have proven to be a promising solution to automate pollen classification. This classification model has shown great effectiveness by accurately classifying and counting the number of pollen grains in microscopic images [27]. For example, a study using five pre-existing neural networks on honey pollen found the InceptionV3 network (a type of CNN) achieved an accuracy of 98.15% [19]. Another comprehensive dataset of almost 19,000 images from 16 pollen types, related to Spanish citrus and rosemary honey, was

Figure 1: Representative pollen grains from the classification training dataset showing morphological diversity across South African taxa which includes 77 distinct pollen types with varying sizes, shapes, and textures captured at different focal depths.

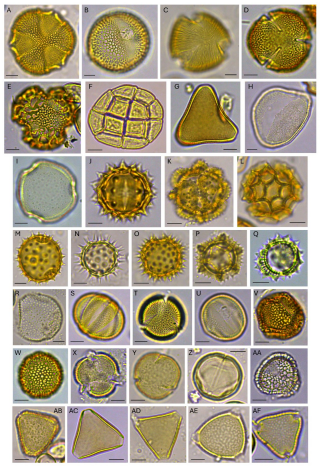

Figure 2: Part of a reference atlas of South African pollen taxa included in the classification dataset. Some from the 77 distinct pollen types show the morphological complexity handled by the POL-ID system.

constructed and used to test various CNNs [12]. In this study, the InceptionV3 network achieved the best accuracy mean of 97.99%.

There has still been a significant challenge for deep learning techniques which is the lack of a substantial labelled dataset required for training [11]. Overfitting is also a concern as validation based on splitting experiments may overestimate the accuracy achievable under new conditions [27]. The millions of parameters in deep networks can also contribute to computational complexity and extensive memory usage which makes them impractical for implementation, as seen in networks like InceptionV3 and DenseNet201 [12]. To address these limitations, a new, simpler network called PolleNetV1 was proposed which achieved 96% accuracy with lower computational and memory effort. Strong visual resemblances between different pollen classes and blurriness in images can hinder classification accuracy [22].

2.2.4 Clustering Methods. When combined with Linear Discriminant Analysis (LDA) and pollen count, PCA proved useful in classifying honey samples by botanical origin, with some types achieving 100% correct classification [4]. However, distinct separation was not always observed, for instance, with Myrtaceae and Salix samples, which showed overlapping due to their diverse pollen sources. Hierarchical Cluster Analysis (HCA) can further refine sample clusters, and in one study, combined with FTIR data and deep learning, it achieved a 96.15% accuracy for clustering honey samples [1].

3 MATERIALS AND METHODS

3.1 Dataset Development and Preparation

3.1.1 Detection Dataset. The detection dataset was built in collaboration with UCT's Department of Chemistry which consisted of microscope slide images of honey samples, each with multiple pollen grains. The dataset included images captured from the same honey samples used for classification. Annotation was carried out in CVAT (Computer Vision Annotation Tool) and exported in YOLO format. The final collection contained thousands of annotations, and covered grains of different sizes, orientations, and focal conditions. Figure 1 shows representative examples of the morphological diversity captured in the training dataset. The dataset was divided at the image level into training, validation, and test sets, using a ratio of 70% for training, 15% for validation, and 15% for testing. Annotations were exported in YOLO format, with each image linked to a text file containing the normalized bounding box coordinates and a single class label representing pollen.

3.1.2 Classification Dataset. The classification dataset images contained pollen grains from honey samples collected specifically for model training. Each image showed one grain of a single known taxon. To capture natural variation in morphology, multiple images of the same grain were taken at different focal depths and magnifications. This allowed the model to learn from the diversity present in individual views.

Following initial model development and evaluation, the dataset was expanded from 75 to 77 taxa through the addition of two previously underrepresented pollen types and supplementary samples for taxa with limited training data (Figure 2 and additionally Figure 10 (Appendix B)). From these, 6588 grain images were extracted for training. Class sizes ranged from as few as 24 to as many as 658 samples per taxon, with a median of 68. This imbalance reflected the natural abundance patterns of pollen types in the samples, where dominant nectar sources such as Lobostemon were represented by hundreds of grains, while rare taxa contributed fewer than 30 samples. While this expanded dataset enabled comparisons (as reported in Section 4.2), attempts to retrain the classification model on the 77-taxa configuration resulted in deteriorated performance on previously well-classified taxa.

3.1.3 Pipeline Validation Dataset. For end-to-end pipeline validation and honey authentication testing, an independent set of nine commercial honey samples was obtained. Six were sourced from different regions across South Africa, two were unknown samples and one was from Angola. The validation set was kept separate from model training and development to ensure an unbiased evaluation.

3.1.4 Data Annotation and Quality Control. Taxonomic identification was carried out by trained palynologist Janais Delport. A custom preprocessing pipeline was developed to extract annotated pollen grains from the original slide images. Each grain was cropped with a standardized procedure that included a ten-pixel border around the bounding box to preserve contextual information. The extracted grains were then resized to 224 by 224 pixels using bicubic interpolation and saved into taxonomically labelled directory structures compatible with deep learning frameworks.

3.1.5 Data Preparation and Augmentation. To prevent data leakage and ensure robust model evaluation, dataset splitting was performed at the grain level rather than the image level, since multiple images of the same grain were often captured at different focal depths and magnifications. All variants of a single grain appeared only in the training, validation, or test set. Stratified sampling was applied to preserve class balance across splits, although this was less effective for the rarest taxa with fewer than 30 examples [3].

Online data augmentation was applied during training using PyTorch's torchvision.transforms library [33]. The augmentation pipeline included geometric transformations, such as random flips and full rotations, to account for arbitrary orientations under the microscope. There were photometric adjustments of brightness, contrast, saturation, and hue to simulate variable illumination. Spatial transformations such as random cropping and limited affine translations (±10% translation, no rotation or shearing) were performed to improve spatial invariance. All training images were normalized with ImageNet statistics to support the use of pretrained weights. Validation and test sets were processed only with resizing and normalization to ensure consistency in evaluation.

3.2 Pollen Grain Detection Module

YOLOv11 was selected as the object detection framework for localizing pollen grains within honey slide images. YOLO architectures are well suited to this task with their single step detection [13]. The nano variant (yolo11n.pt) was employed to make the approach more feasible for resource-constrained settings [37].

The model was initialized with COCO (Common Objects in Context) pre-trained weights to take advantage of low-level feature representations such as edge and shape detection. This transfers well to the circular and elliptical morphologies of pollen grains [34]. Hyperparameters included an input resolution of 640 by 640 pixels, a batch size of 16, and a maximum of 300 epochs with early stopping after 30 epochs of no improvement. Online data augmentation was applied during training to improve generalization across different imaging conditions.

Performance was evaluated using standard object detection metrics. The primary measure was mAP@0.5, with additional reporting of mAP across IoU thresholds from 0.5 to 0.95. Precision and recall curves were used to analyse trade-offs between false positives and false negatives.

3.3 CNN-Based Pollen Classification Module

3.3.1 Model Architecture Selection. ConvNeXt was selected for pollen grain classification because of its strength in fine-grained recognition and its balance of accuracy and efficiency. This approach is well suited to pollen taxonomy, where subtle differences between morphologically similar grains must be distinguished [16]. The ConvNeXt Tiny variant was used as the baseline model. To explore the impact of model capacity, ConvNeXt Small, and other architectures like EfficientNet and ResNet, were also evaluated to test whether larger architectures could improve accuracy.

3.3.2 Transfer Learning Strategy. All models were initialized with ImageNet pre-trained weights. This provided a foundation of low-level feature representations such as edges, textures, and shapes, which transfer well to pollen microscopy [40]. The final classification layer was replaced with a new head matching the 77 pollen taxa in the dataset. Fine-tuning the feature extractors allowed the models to adapt to the domain-specific requirements of pollen classification [31]. It leveraged the representational strength of pretraining to compensate for the limited dataset size.

3.3.3 Training Configuration and Optimization. Models were trained using the AdamW optimizer with an initial learning rate of 8×10^{-5} and a weight decay of 0.015 [20]. A ReduceLROnPlateau scheduler dynamically lowered the learning rate by a factor of 0.7 after six stagnant epochs and enabled stable convergence while avoiding overfitting [28]. Training ran for a maximum of 45 epochs, with early stopping triggered when validation accuracy failed to improve. A batch size of 20 was chosen to balance memory use with stable gradient estimates.

3.3.4 Loss Function and Class Imbalance Handling. The dataset showed strong natural imbalance, with class counts ranging from fewer than 30 to more than 600 grains per taxon. This was addressed with stratified sampling across the 70-15-15 split, though the smallest classes remained underrepresented. During training, minority taxa with fewer than 50 examples were upweighted (up to 3×) in the loss function to improve their recognition without harming performance on dominant classes. To improve generalization, label smoothing cross-entropy with $\varepsilon = 0.1$ was employed, which reduced overconfidence in predictions and mitigated label noise [21]. This was especially important where taxonomic boundaries were subtle.

3.3.5 Model Evaluation and Validation. Performance was assessed using multiple metrics. Accuracy provided an overall measure of performance, while macro-averaged precision, recall, and F1-scores accounted for class imbalance. Per-class results were analysed to highlight taxa with persistent misclassifications. Model stability was further tested through repeated runs with different random seeds, providing cross-validation of performance. Experimental conditions were compared statistically.

3.3.6 Experimental Conditions and Ablation Studies. Several experimental configurations were designed to explore the effect of different training strategies on classification performance. The baseline configuration used ConvNeXt Tiny with standard cross-entropy loss and basic augmentation, providing a reference point for subsequent experiments. Additional experiments introduced enhanced regularization through label smoothing, increased dropout, the impact of focal loss, cosine annealing learning rate schedules, gradient clipping, ensembles and test-time augmentation.

3.4 Novel Pollen Type Discovery Through Clustering

To explore potential novel pollen types beyond the supervised classification labels, deep feature vectors were extracted from the penultimate layer of the trained ConvNeXt model. These 768-dimensional embeddings capture high-level morphological patterns and discard low-level noise. Features from all available datasets (training, validation, and test) were pooled to maximize morphological diversity and then standardized via z-score normalization for unsupervised analysis.

Dimensionality reduction was applied prior to clustering: PCA was used to denoise features and retained 95% variance, and UMAP provided reduced spaces (10D for clustering, 2D for visualization) that preserved local structure more effectively than linear projections. HDBSCAN was selected as the primary clustering algorithm due to its ability to handle variable densities and identify noise points that may represent rare or novel taxa [23]. Parameter sweeps over minimum cluster size, minimum samples, and distance metrics were conducted, with evaluation based on clustering validity indices.

Cluster purity was assessed relative to taxonomic labels. Homogeneous clusters validated model embeddings, while mixed clusters flagged possible mislabels. Noise points and local outliers were examined separately through a protocol which combined visual inspection, reference comparison and contextual ecological information. While CNN-based clustering provides a useful exploratory tool for highlighting ambiguous cases, its effectiveness is limited by biases in the training data [7].

3.5 Honey Authentication Pipeline

The POL-ID system, as shown in Figure 3, processes raw microscopic honey slide images into standardized authentication reports that follow International Commission for Bee Botany (ICBB) protocols. The workflow has four components: (1) pollen grain detection with YOLO, (2) taxonomic classification with ConvNeXt, (3) novel type discovery with HDBSCAN clustering, and (4) honey authentication based on melissopalynological standards. Each slide is processed step by step with built-in error checks to handle variation in image quality and honey type.

Pollen grains are first located with YOLO. Inputs were standardized consistent with model training. Classification outputs are routed based on confidence. High-confidence grains ($\geq$ 70%) are assigned labels directly. Low-confidence grains ($<$ 70%) bypass forced classification and instead enter the clustering stage, where features are extracted from the CNN's penultimate layer to capture morphological patterns.

Low-confidence grains are pooled and clustered using HDBSCAN. Single-grain outliers are flagged for expert review, while clusters with multiple samples are treated as novel candidates. The pipeline applies established melissopalynological thresholds [35]. This ensures results are aligned with international norms for honey authentication. Final reports give the honey classification, dominant taxa, confidence scores, full pollen counts, grain-level results, and clustering outcomes. Samples with too many low-confidence results are flagged for review to avoid misinterpretation.

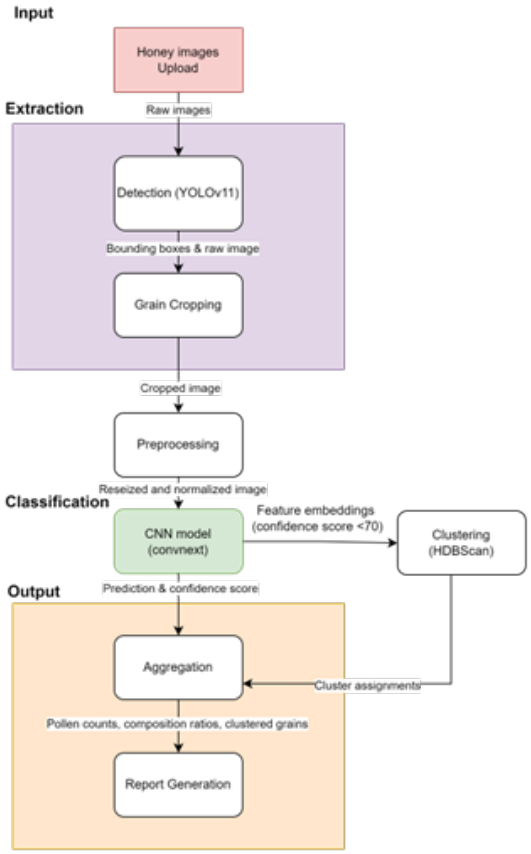

Figure 3: POL-ID system flowchart showing the complete pipeline from honey image upload through detection, classification, and clustering to final report generation.

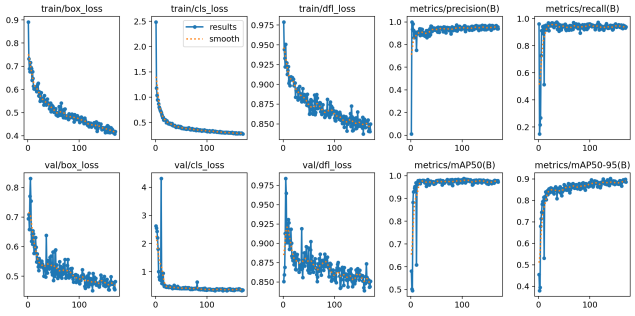

Figure 4: YOLOv11-nano training performance showing convergence of loss functions and detection metrics over 300 epochs.

4 RESULTS

4.1 Detection Module Performance

The YOLOv11-nano detector achieved strong performance in identifying pollen grains on honey slides. On the test set, it reached

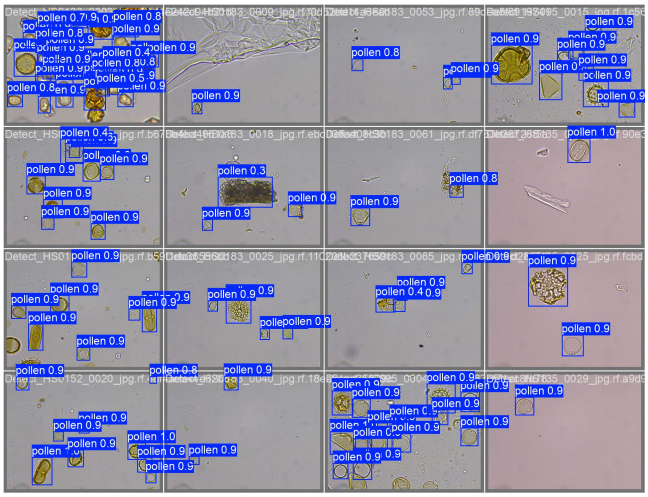

Figure 5: Detection results showing YOLO bounding boxes with confidence scores on honey slide images. The system identifies pollen grains across different morphologies and imaging conditions.

a mean Average Precision of 97.73% at IoU 0.5, above the initial 80-85% target for reliable authentication. Performance remained robust across stricter thresholds, with mAP@0.5–0.95 of 91.29%.

Precision (96.36%) and recall (96.58%) showed that the system consistently identified true pollen grains with few false positives or missed detections. Of 211 annotated grains, 195 were correctly detected, with only 8 false positives and 16 false negatives. The detector showed no false positive detections on background regions, preventing debris or air bubbles from being miscounted as pollen.

Training converged over 300 epochs, with both training and validation losses following similar trajectories (Figure 4). Box regression, classification, and focal loss values all decreased smoothly, and performance plateaued around epoch 100. Precision and recall stabilised above 0.95 early in training.

With 2.59 million parameters and 6.4 GFLOPs, the nano variant processed 640 × 640 images quickly without sacrificing accuracy. Detection examples demonstrate the system's ability to accurately localize pollen grains with high confidence scores (Figure 5). The confidence values range from 0.3 to 1.0. Overall, the detector met its design goals: high accuracy and low false detections.

4.2 CNN Classification Performance

The ConvNeXt-based classification module demonstrated strong and reproducible performance across multiple experimental settings. Test accuracies ranged from 89.04% to 94.43%, depending on architecture, loss function, augmentation strategy, and training protocol. These results establish ConvNeXt Tiny as a baseline for South African pollen classification while revealing the limits of more complex techniques.

4.2.1 Baseline and Targeted Improvements. The original ConvNeXt Tiny baseline achieved 93.51% test accuracy (Figure 6) with balanced precision, recall, and F1-scores above 93%, surpassing the project's 80-85% target. Conservative refinements produced the best overall

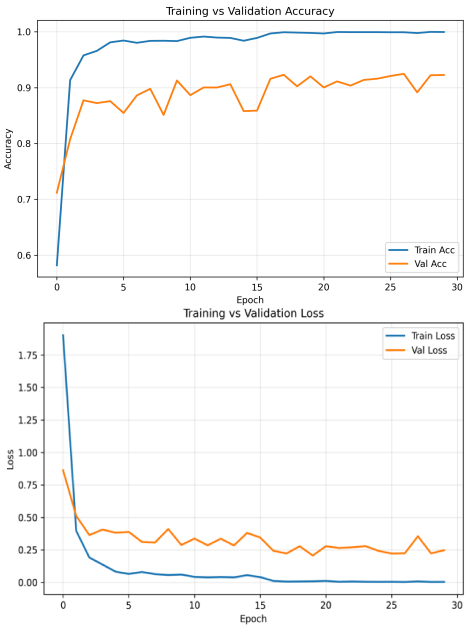

Figure 6: Classification performance of Convnext-tiny

performance: a targeted ConvNeXt configuration with an enhanced classifier, label smoothing, and moderate augmentation reached a peak of 94.43% test accuracy and 94.97% validation accuracy. This tied the best result later obtained by a larger ConvNeXt Base model.

4.2.2 Comparative Model and Strategy Analysis. Comparative testing highlighted distinct performance profiles across CNN families and testing strategies (Table 1), a visualisation is shown in Figure 11 (Appendix B) and an expanded table is shown in Table 5 (Appendix A). ConvNeXt Tiny outperformed both ResNet50 and EfficientNet-B3, with diminishing returns when scaling to ConvNeXt Small or Base. Focal loss, often recommended for class imbalance, consistently underperformed: across all architectures it reduced F1-scores and occasionally collapsed small classes (e.g., Daisy_sp2). Similarly, overly aggressive augmentation strategies (MixUp, CutMix, Gaussian blur, heavy erasing) degraded accuracy to 91–92%. By contrast, test-time augmentation (TTA) produced modest but stable improvements. The best configuration, Targeted ConvNeXt with full TTA, achieved 94.66% test accuracy, a +1.15 percentage point gain over the baseline.

Table 1: Comparative CNN model results for pollen classification.

Model / Strategy	Test Acc	Test F1	Val Acc
ConvNeXt Tiny (Baseline)	93.51%	93.61%	92.48%
Targeted ConvNeXt (Tiny)	94.43%	91.58%	94.97%
ConvNeXt Base (Advanced)	94.43%	93.20%	92.51%
ResNet50 + Focal Loss	92.11%	90.90%	90.15%
EfficientNet-B3 + Focal Loss	89.04%	88.26%	89.73%

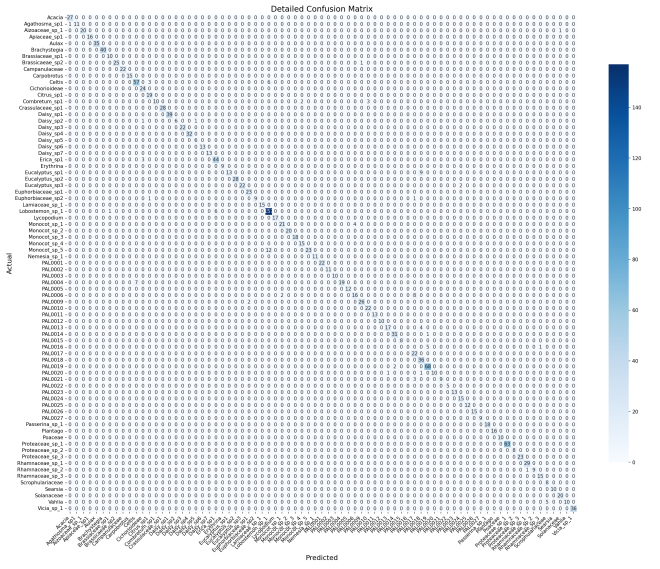

Figure 7: Confusion matrix analysis showing the misclassifications between similar taxa.

4.2.3 Per-Class Performance, Confusion Patterns, and Persistent Challenges. While overall metrics were strong, class-level analysis revealed persistent weaknesses. Performance varied widely across taxa due to class imbalance and morphological similarity. Many abundant taxa achieved perfect scores, including Apiaceae_sp1, Aulax, Crassulaceae_sp1, Poaceae, and several Daisy species. However, extremely rare taxa consistently failed: PAL0016 (11 samples) achieved 0% accuracy across all experiments, and Daisy_sp2 (8 samples) oscillated between complete failure (0%) and partial recovery (62.5% recall). Accuracy strongly correlated with sample size: taxa with 50 training samples frequently exceeded 90% accuracy, while those with 20 often dropped below 70%. Confusion matrix analysis (Figure 7) highlighted systematic misclassifications between morphologically similar taxa, particularly Daisy_sp3 vs Daisy_sp4 and PAL0016 vs PAL0014, with error rates exceeding 65–80%.

4.2.4 Summary. The classifier generally produced confident predictions, with over 90% of outputs exceeding 90% confidence. However, misclassified samples often retained relatively high confidence (0.72-0.90), showing that confidence alone does not guarantee correctness. Weighted sampling, label smoothing, and focal loss provided partial mitigation for class imbalance, but rare taxa remained difficult to classify consistently. Model accuracy showed a strong dependence on sample size, with taxa represented by fewer than 20 grains often falling below 70% accuracy, while those with more than 50 grains exceeded 90%.Confusion matrix analysis (Figure 8) revealed systematic misclassifications between morphologically similar taxa, especially within the Daisy group and among PAL codes. These errors were consistent across models and reflected genuine morphological overlap.

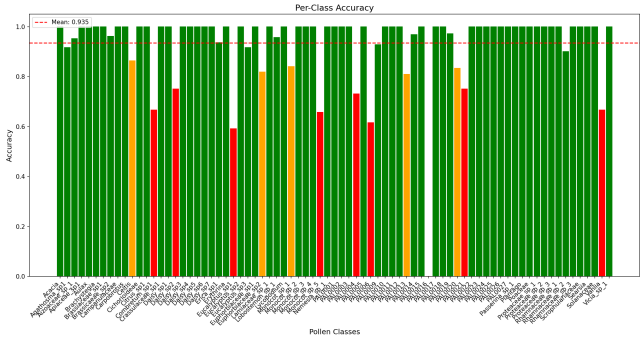

Figure 8: Per-class acuuracy showing the misclassifications of specific pollen types.

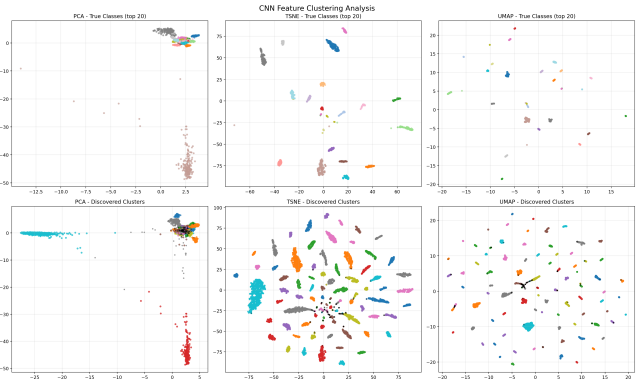

Figure 9: CNN feature clustering analysis using different dimensionality reduction techniques. Top row shows clustering based on true taxonomic classes (top 20), while bottom row shows HDBSCAN-discovered clusters.

4.3 Clustering Module Effectiveness

The HDBSCAN-based clustering module demonstrated strong potential for validating supervised classification patterns and exploring novel pollen types through unsupervised analysis of CNN feature embeddings. Operating on high-dimensional feature vectors from the ConvNeXt model, the clustering pipeline achieved an Adjusted Rand Index (ARI) of 0.962, indicating close agreement with expert-assigned taxonomic labels.

Figure 9 illustrates the effectiveness of different dimensionality reduction techniques for visualizing pollen grain relationships. With a minimum cluster size of 10 and 10 minimum samples, the algorithm produced 78 distinct clusters from more than 7,600 grains, assigning only 1.8% as noise. Most clusters were highly pure, with 77 out of 78 matching a single pollen type at over 90%. Notably, taxa that were difficult for supervised classification formed clean and homogeneous clusters. The 136 noise samples identified by clustering represent promising candidates for novel pollen type discovery.

4.4 End-to-End Honey Authentication Results

The POL-ID pipeline was validated on eight independent South African honey samples and one honey sample from Angola. In total, 1,307 slides and 3,969 pollen grains were processed (Table 2). Each sample was classified according to ICBB thresholds, and results were compared with manual melissopalynological analysis.

4.4.1 Pipeline Model Configuration. The end-to-end POL-ID pipeline integrates three core components: YOLOv11-nano for pollen grain detection (achieving 97.73% mAP@0.5), the baseline ConvNeXt Tiny for species classification (achieving 92.96% accuracy) , and HDBSCAN clustering for novel pollen type discovery (ARI = 0.883). Due to technical constraints encountered during model retraining, the pipeline validation employs a ConvNeXt model and clustering module trained on 75 pollen taxa, while the architectural comparisons in Section 4.2 utilized models trained on an expanded 77-taxa dataset. The detection module maintain consistent performance across both configurations as it operates independently of the specific number of classification taxa.

4.4.2 Pipeline Performance Overview. The automated system processed each sample in approximately 5-10 minutes. All nine samples were assigned classifications consistent with ICBB standards, including monofloral, mixed, and multifloral honeys.

Table 2: Overview of honey sample processing and classification

Sample	Slides	Grains	Classification
HS095	33	307	Monofloral (Celtis)
HS133	41	451	Monofloral (Lobostemon sp. 1)
HS135	192	351	Monofloral (PAL0019)
HS150	68	538	Multifloral
HS152	235	580	Mixed (Apiaceae sp. 1)
HS170	152	620	Multifloral
HS177	214	418	Multifloral
HS183	182	361	Multifloral
HS189	190	343	Multifloral
Total	1,307	3,969	–

4.4.3 Monofloral Honey Authentication. Three samples were classified as monofloral. The detailed comparison between automated POL-ID analysis and manual melissopalynological analysis is shown in Table 3.

4.4.4 Multifloral Honey Recognition. Three samples were classified as multifloral. The comparison between automated and manual analysis is presented in Table 4.

4.4.5 Uncertainty Handling and Novel Type Detection. Across all samples, between 8.4% and 21.7% of grains were assigned to uncertain categories. These included low-confidence CNN predictions and clusters not matched to existing taxa.

4.4.6 Comparative Results with Manual Analysis. Automated classification matched manual assessments of honey type in all nine samples. However, quantitative analysis revealed systematic differences in pollen abundance estimates. For monofloral honeys, POL-ID consistently underestimated dominant pollen percentages compared to manual analysis, with differences ranging from -9.5% (Celtis) to -22.9% (Lobostemon sp. 1). The mixed honey sample (HS152) was correctly classified with Apiaceae sp. 1 as the predominant taxon, but POL-ID estimated 23.6% abundance versus 44%

Table 3: Monofloral honey authentication results: POL-ID vs manual analysis

Sample	Taxon	POL-ID (%)	Manual (%)	Diff. (%)
HS095 (Celtis)	Celtis	46.9	56.4	-9.5
	Proteaceae sp. 1	10.7	14.3	-3.6
	Campanulac.	9.1	8.3	+0.8
HS135 (PAL0019)	PAL0019	45.3	56.3	-11.0
	PAL0018	11.1	16.8	-5.7
	Brachystegia	6.8	18.2	-11.4
HS133 (Lobostemon)	Lobostemon sp. 1	55.4	78.3	-22.9
	Erica sp. 1	5.8	9.7	-3.9
	Vicia sp. 1	0.7	3.5	-2.8

Table 4: Multifloral honey recognition results: dominant taxa comparison

Sample	Dominant Taxon (POL-ID)	POL-ID (%)	Manual Top Taxa	Manual (%)
HS150	Lobostemon sp. 1	19.1	Brassica. sp. 2; PAL0010; Apiaceae sp. 1	19.6; 17.0; 11.2
HS170	PAL0013	11.5	Vahlia-type sp. 1; Lobostemon sp. 1	20.0; 14.8
HS177	Eucalyptus sp. 1	12.0	Eucalyptus sp. 3; Eucalyptus sp. 2	32.2; 25.4
HS183	Lobostemon sp. 1	14.7	Lobostemon sp. 1	27.5
HS189	PAL0011	16.3	PAL0011; Rhamnaceae sp. 1; Multiple taxa tie	29.0; 21.6; 7.1

in manual analysis. Multifloral honey classifications showed variable agreement in dominant taxa identification, with some samples (HS183) showing good correspondence and others (HS150, HS189) identifying different primary taxa than manual analysis. Despite these quantitative discrepancies, all samples received correct ICBB-compliant honey type classifications. Automated analysis required 5-10 minutes per sample compared to 180-300 minutes manually, and repeated runs produced consistent outputs.

5 DISCUSSION

The findings of POL-ID show that automated pollen analysis can not only complement but in many cases improve on traditional manual methods, especially when applied to the complex and highly diverse flora of South Africa.

5.1 Technical Performance and Methodological Advances

5.1.1 Detection Module Achievements. The YOLOv11-nano detector achieved an mAP@0.5 of 97.73%, which is a substantial leap from earlier pollen detection systems. For instance, He et al. [11] reported precision of 0.663 and recall of 0.914 using YOLOv2, while our system reached precision of 96.36% and recall of 96.58%. This improvement can be attributed both to advances in YOLO architecture and to the carefully curated dataset, which included the varied imaging conditions and morphologies typical of South African honey. The detector's ability to correctly identify background noise such as debris and air bubbles, a recurring obstacle in microscopic analysis, is important. Starting from COCO pre-trained weights allowed the model to adapt natural-image features to pollen shapes, while the lightweight YOLOv11-nano architecture ensures fast processing for laboratory use.

5.1.2 Classification Performance and Architectural Choices. The ConvNeXt-based classifier achieved 94.43% accuracy across 77 pollen taxa. Although some studies report higher accuracies with 97.99% on 16 Spanish taxa [12], or 98.15% on honey pollen [19], our system worked with a completely new dataset and exceeded our original research aims of obtaining an accuracy of between 80–85%. ConvNeXt Tiny consistently outperformed not only ResNet50 and EfficientNet-B3 but also its larger sibling, ConvNeXt Small, achieving comparable results to ConvNeXt Base despite having significantly fewer parameters. This result supports recent findings that architectural refinements often matter more than simply scaling up model size [27]. Contrary to expectations, focal loss consistently underperformed across all architectures tested. Standard cross-entropy and label smoothing provided more reliable results for this dataset.

5.1.3 Clustering Innovation for Uncertainty Quantification. HDBSCAN clustering achieved an Adjusted Rand Index (ARI) of 0.962, showing strong alignment between CNN-derived features and expert taxonomy. This result highlights that the CNN learned features with genuine morphological meaning. Unlike many past systems, POL-ID incorporates uncertainty quantification through systematic noise detection, identifying 136 samples (1.8%) that fall outside known taxonomic categories. The clustering analysis produced 78 distinct clusters from over 7,600 pollen grains, with 77 of 78 clusters exceeding 90% taxonomic purity. Interestingly, some poorly classified taxa (e.g., PAL0016) still formed coherent clusters. This suggests that limited training data, rather than a failure of feature learning, accounts for weaker supervised performance. Such insights point the way towards more targeted data collection in future work.

5.2 Honey Authentication Performance and Practical Validation

5.2.1 ICBB Compliance and Classification Accuracy. POL-ID successfully classified all nine honey samples according to ICBB standards. This included monofloral honeys, mixed honeys, and complex multifloral samples from different regions across South Africa and Angola. The system correctly identified honey types even when automated and manual pollen counts differed substantially (for example, HS152, where Apiaceae representation was 23.6% versus 44.0%, or HS133 where Lobostemon was 55.4% versus 78.3% manually). The model underestimated dominant pollen percentages, which could cause problems in borderline cases. Although the tested samples had enough margin for correct classification, honeys with dominant taxa close to ICBB thresholds risk misclassification. This was most evident in mixed honeys, where POL-ID often missed the correct dominant taxon even if the overall honey type was right. For commercial authentication, borderline cases would still need manual checks to confirm botanical origin.

5.2.2 Processing Efficiency and Scalability. The system cut processing time by an order of magnitude, reducing analysis from 180-300 minutes manually to just 15-20 minutes. This efficiency with reproducibility across repeated runs, addresses the weaknesses of manual analysis- subjectivity and inconsistency. In this study, POL-ID processed 1,307 slides and analysed 3,969 grains, demonstrating scalability well beyond what would be feasible by hand. If applied at scale, such improvements could support authentication across the country's annual honey production.

5.3 Limitations and Technical Challenges

5.3.1 Morphological Convergence and Taxonomic Resolution. Confusion between morphologically similar taxa reflects the limits of image-based classification. The most problematic misclassifications included PAL0016 vs PAL0019 (83.3% error rate), Monocot_sp_5 vs Lobostemon_sp_1 (34.3% error rate), and Eucalyptus_sp1 vs PAL0018 (40.9% error rate). These are not failures unique to machine learning. Expert palynologists face the same difficulties [10]. For genera such as Eucalyptus, where pollen grains are notoriously uniform [26], species-level identification may ultimately require chemical or genetic data. The high error rates between taxonomically distinct groups (e.g., Monocot_sp_5 and Lobostemon_sp_1) suggest that morphological convergence extends beyond closely related taxa. There is fundamental challenges of pollen identification based solely on visual features.

5.3.2 Class Imbalance and Data Scarcity. Classification accuracy correlated strongly with training sample size. Taxa with more than 100 training samples, such as Lobostemon_sp_1, Celtis, and Monocot_sp_5, generally performed well, though even large classes were not immune to errors (Monocot_sp_5 reached only 65.7% accuracy). In contrast, underrepresented taxa consistently failed:

PAL0016 with 11 samples scored 0%, while Daisy_sp2 (8 samples) and PAL0022 (12 samples) showed erratic results. Weighted sampling and focal loss helped to some degree, but the lack of data cannot be fully offset by algorithms. Reliable performance appeared to require at least 40-50 examples. Rare taxa will therefore need targeted data collection, likely in collaboration with botanical institutions.

5.3.3 Dataset Expansion Challenges and Model Stability. Expanding the model from 75 to 77 taxa exposed a key limitation of deep learning. Although the original classes were unchanged and training accuracy remained high, the updated model performed worse in the pipeline, with reliable taxa such as Celtis and Poaceae now misclassified. This form of catastrophic forgetting arises when adding new classes reshapes the feature space and disrupts earlier decision boundaries. The stability of the 75-class model highlights the need to build balanced datasets from the start rather than adding classes incrementally. For real-world use, this means updates must involve careful rebalancing of the dataset. Future work should explore continual learning methods that expand taxonomic coverage without sacrificing existing performance.

5.3.4 Conservative Classification and Threshold Effects. The chosen 70% confidence threshold ensured high precision but left 8–22% of grains marked as uncertain. Relaxing this threshold could improve quantitative accuracy, but the conservative setting has clear advantages. It prevents overconfident misclassification and provides transparency, allowing human experts to step in where needed.

5.4 Implications for South Africa and Future Directions

5.4.1 South Afircan Pollen Landscape Implications. South Africa's flora creates particular challenges for honey authentication. Windblown pollen and the coexistence of diverse pollen complicate traditional manual methods [26]. POL-ID was built with these conditions in mind, trained specifically on regional taxa and designed to flag uncertain grains. The system's adaptability through clustering means it can accommodate new discoveries.

The economic and regulatory implications are equally significant. Honey fraud is a global issue, costing producers hundreds of millions of dollars annually [42]. Automated, standardised authentication could help South African producers secure premium markets, especially for distinctive honeys such as those from fynbos regions. The ICBB-compliant output format ensures compatibility with international requirements, while reduced reliance on specialist palynologists opens access to smaller-scale producers who may previously have been excluded by cost.

Finally, the framework developed here has broader relevance. The integration of detection, fine-grained classification, and uncertainty quantification can be applied to other plant-based food products where authenticity matters. If the system can handle South Africa's extreme taxonomic diversity, it is likely adaptable to other biodiversity hotspots.

5.4.2 Future Directions. System performance could be improved by integrating additional morphological descriptors such as size and surface texture, or by combining image-based methods with pollen chemistry or DNA barcoding. Ensemble architectures may further increase robustness. Active learning could guide efficient expansion of training datasets by highlighting the most informative new samples.

To sustain progress, a comprehensive South African pollen image database is needed. Such a resource could be developed in collaboration with universities and beekeeping associations. At the international level, standardised databases and analysis protocols could support comparative studies and expand the reach of authenticated honey into global markets.

6 CONCLUSIONS

This work developed POL-ID, the first automated honey authentication system tailored to South African pollen taxa. The system integrates YOLOv11-nano object detection (97.73% mAP@0.5), ConvNeXt-based classification (94.66% accuracy with test-time augmentation), and HDBSCAN clustering (ARI = 0.962) to handle melissopalynological complexity while maintaining ICBB compliance. Validation on nine independent honey samples demonstrated correct classification of all honey types, though systematic underestimation of dominant pollen percentages revealed limitations for borderline cases near classification thresholds.

Systematic evaluation across different experimental configurations established that architectural refinements outperform complex training strategies, with ConvNeXt consistently superior to ResNet and EfficientNet architectures. Contrary to expectations, focal loss degraded performance across all tested configurations, while standard cross-entropy with conservative augmentation proved most effective for pollen classification.

Applying the pipeline to South African honeys, which draw on one of the world's richest floras, showed that the system can manage extreme taxonomic diversity. The approach is scalable, meaning it could be extended to other regions facing similar challenges. The system successfully manages South Africa's exceptional taxonomic diversity, demonstrating scalability for biodiversity hotspots where traditional authentication struggles. Processing efficiency improved while maintaining reproducibility. However, morphological convergence between distant taxa and failures on severely underrepresented classes highlight fundamental limitations requiring targeted data collection strategies. Future work could integrate additional data sources to refine species-level identification. Broader datasets and active learning strategies could also improve coverage of rare taxa.

In summary, POL-ID shows that automated systems can meet the practical and scientific demands of honey authentication in a biodiversity hotspot. It demonstrates a path forward where AI augments traditional palynology and can provide automated authentication.

REFERENCES

[1] Fatih Mehmet Avcu. 2024. Clustering honey samples with unsupervised machine learning methods using FTIR data. *Anais da Academia Brasileira de Ciências* 96, 1 (2024). DOI:https://doi.org/10.1590/0001-3765202420230409

[2] Ricardo J. G. B. Campello, Davoud Moulavi, and Joerg Sander. 2013. Density-Based Clustering Based on Hierarchical Density Estimates. Advances in Knowledge Discovery and Data Mining 7819, (2013), 160–172. DOI:https://doi.org/10.1007/978-3-642-37456-2_14

[3] Davide Chicco. 2017. Ten quick tips for machine learning in computational biology. BioData Mining 10, 1 (December 2017). DOI:https://doi.org/10.1186/s13040-017-0155-3

[4] Eduardo Corbella and Daniel Cozzolino. 2008. Combining Multivariate Analysis and Pollen Count to Classify Honey Samples Accordingly to Different Botanical Origins. *Chilean Journal of Agricultural Research* 68, 1 (March 2008). DOI:https://doi.org/10.4067/s0718-58392008000100010

[5] Benoît Crouzy, Michelle Stella, Thomas Konzelmann, Bertrand Calpini, and Bernard Clot. 2016. All-optical automatic pollen identification: Towards an operational system. *Atmospheric Environment* 140, (September 2016), 202–212. DOI:https://doi.org/10.1016/j.atmosenv.2016.05.062

[6] David L. Davies and Donald W. Bouldin. 1979. A Cluster Separation Measure. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-1, 2 (April 1979), 224–227. DOI:https://doi.org/10.1109/TPAMI.1979.4766909

[7] Ekberjan Derman. 2021. Dataset Bias Mitigation Through Analysis of CNN Training Scores. arXiv (Cornell University) (January 2021). DOI:https://doi.org/10.48550/arxiv.2106.14829

[8] Dilpreet Singh Brar, Ashwani Kumar Aggarwal, Vikas Nanda, Sudhanshu Saxena, and Satyendra Gautam. 2023. AI and CV based 2D-CNN Algorithm: Botanical Authentication of Indian Honey Varieties. *Sustainable Food Technology* (January 2023), 373–385. DOI:https://doi.org/10.1039/d3fb00170a

[9] Susanne Dunker, Matthew Boyd, Walter Durka, Silvio Erler, W Stanley Harpole, Silvia Henning, Ulrike Herzschuh, Thomas Hornick, Tiffany Knight, Stefan Lips, Patrick Mäder, Elena Motivans Švara, Steven Mozarowski, Demetra Rakosy, Christine Römermann, Mechthild Schmitt-Jansen, Kathleen Stoof-Leichsenring, Frank Stratmann, Regina Treudler, and Risto Virtanen. 2022. The potential of multispectral imaging flow cytometry for environmental monitoring. Cytometry Part A 101, 9 (June 2022), 782–799. DOI:https://doi.org/10.1002/cyto.a.24658

[10] Ariadne Barbosa Gonçalves, Junior Silva Souza, Gercina Gonçalves da Silva, Marney Pascoli Cereda, Arnildo Pott, Marco Hiroshi Naka, and Hemerson Pistori. 2016. Feature Extraction and Machine Learning for the Classification of Brazilian Savannah Pollen Grains. *PLOS ONE* 11, 6 (June 2016), e0157044. DOI:https://doi.org/10.1371/journal.pone.0157044

[11] Chloe He, Alexis Gkantiragas, and Gerard Glowacki. 2018. Honey Authentication with Machine Learning Augmented Bright-Field Microscopy. *arXiv preprint arXiv:1901.00516*.

[12] 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 (August 2023), 105605. DOI:https://doi.org/10.1016/j.jfca.2023.105605

[13] Redmon Joseph, Divvala Santosh, Girshick Ross, and Farhadi Ali. 2016. You Only Look Once: Unified, Real-Time Object Detection. arXiv (Cornell University) (January 2016). DOI:https://doi.org/10.48550/arxiv.1506.02640

[14] Y. Kaya, M. E. Erez, O. Karabacak, L. Kayci, and M. Fidan. 2013. An automatic identification method for the comparison of plant and honey pollen based on GLCM texture features and artificial neural network. *Grana* 52, 1 (February 2013), 71–77. DOI:https://doi.org/10.1080/00173134.2012.754050

[15] Elżbieta Kubera, Agnieszka Kubik-Komar, Paweł Kurasiński, Krystyna Piotrowska-Weryszko, and Magdalena Skrzypiec. 2022. Detection and Recognition of Pollen Grains in Multilabel Microscopic Images. Sensors 22, 7 (March 2022), 2690–2690. DOI:https://doi.org/10.3390/s22072690

[16] Zhiheng Li, Tongcheng Gu, Bing Li, Wubin Xu, Xin He, and Xiangyu Hui. 2022. ConvNeXt-Based Fine-Grained Image Classification and Bilinear Attention Mechanism Model. Applied Sciences 12, 18 (September 2022), 9016. DOI:https://doi.org/10.3390/app12189016

[17] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. 2017. Focal Loss for Dense Object Detection. arXiv (Cornell University) (August 2017). DOI:https://doi.org/10.48550/arxiv.1708.02002

[18] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. 2022. A ConvNet for the 2020s. (January 2022). DOI:https://doi.org/10.48550/arxiv.2201.03545

[19] Fernando López-García, José Miguel Valiente-González, Isabel Escriche-Roberto, Marisol Juan-Borrás, Mario Visquert-Fas, Vicente Atienza-Vanacloig, and Manuel Agustí-Melchor. 2023. Classification of Honey Pollens with ImageNet Neural Networks. *Lecture Notes in Computer Science* 14185 (January 2023), 192–200. DOI:https://doi.org/10.1007/978-3-031-44240-7_19

[20] Ilya Loshchilov and Frank Hutter. 2019. Decoupled Weight Decay Regularization. arXiv.org. DOI:https://doi.org/10.48550/arXiv.1711.05101

[21] Michal Lukasik, Srinadh Bhojanapalli, Aditya Krishna Menon, and Sanjiv Kumar. 2020. Does label smoothing mitigate label noise? arXiv (Cornell University) (January 2020). DOI:https://doi.org/10.48550/arxiv.2003.02819

[22] 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:https://doi.org/10.1016/j.jksuci.2023.01.013

[23] Claudia Malzer and Marcus Baum. 2020. A Hybrid Approach To Hierarchical Density-based Cluster Selection. 2020 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI) (September 2020), 223–228. DOI:https://doi.org/10.1109/MFI49285.2020.9235263

[24] Mashudu Patience Mamathaba, Kowiyou Yessoufou, and Annah Moteetee. 2022. What Does It Take to Further Our Knowledge of Plant Diversity in the Megadiverse South Africa? Diversity 14, 9 (September 2022), 748. DOI:https://doi.org/10.3390/d14090748

[25] Leland McInnes, John Healy, and James Melville. 2020. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv.org. DOI:https://doi.org/10.48550/arXiv.1802.03426

[26] Nikiwe Ndlovu, Frank H. Neumann, Michelle D. Henley, Robin M. Cook, and Chevonne Reynolds. 2023. Melissopalynological investigations of seasonal honey samples from the Greater Kruger National Park, Savanna biome of South Africa. *Palynology* (February 2023). DOI:https://doi.org/10.1080/01916122.2023.2179679

[27] 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 (March 2021), 850–862. DOI:https://doi.org/10.1111/2041-210x.13575

[28] PyTorch Contributors. 2025. ReduceLROnPlateau. Pytorch.org. Retrieved September 4, 2025 from https://docs.pytorch.org/docs/stable/generated/torch.optim.lr_scheduler.ReduceLROnPlateau.html?

[29] Peter J. Rousseeuw. 1987. Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis. Journal of Computational and Applied Mathematics 20, 0377-0427 (November 1987), 53–65. DOI:https://doi.org/10.1016/0377-0427(87)90125-7

[30] Connor Shorten and Taghi M. Khoshgoftaar. 2019. A survey on Image Data Augmentation for Deep Learning. Journal of Big Data 6, 1 (July 2019). DOI:https://doi.org/10.1186/s40537-019-0197-0

[31] Nima Tajbakhsh, Jae Y. Shin, Suryakanth R. Gurudu, R. Todd Hurst, Christopher B. Kendall, Michael B. Gotway, and Jianming Liang. 2016. Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning? IEEE Transactions on Medical Imaging 35, 5 (May 2016), 1299–1312. DOI:https://doi.org/10.1109/tmi.2016.2535302

[32] Ofijan Tesfaye, Asnake Desalegn, and Diriba Muleta. 2024. Melissopalynological analysis and microbiological safety of fresh and market honey (Apis mellifera L. and Meliponula beccarii L.) from Western Oromia, Ethiopia. Heliyon 10, 7 (April 2024), e28185–e28185. DOI:https://doi.org/10.1016/j.heliyon.2024.e2818

[33] Wei-Ming Thor. 2025. Data Transformations (‘torchvision.transforms‘). Apxml.com. Retrieved September 4, 2025 from https://apxml.com/courses/getting-started-with-pytorch/chapter-5-efficient-data-handling/data-transformations-torchvision-transforms?

[34] Nikos Tsiknakis, Elisavet Savvidaki, Georgios C. Manikis, Panagiota Gotsiou, Ilektra Remoundou, Kostas Marias, Eleftherios Alissandrakis, and Nikolas Vidakis. 2022. Pollen Grain Classification Based on Ensemble Transfer Learning on the Cretan Pollen Dataset. Plants 11, 7 (March 2022), 919. DOI:https://doi.org/10.3390/plants11070919

[35] 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:https://doi.org/10.1051/apido:2004050

[36] Matthijs J. Warrens and Hanneke van der Hoef. 2022. Understanding the Adjusted Rand Index and Other Partition Comparison Indices Based on Counting Object Pairs. Journal of Classification 39, 3 (July 2022), 487–509. DOI:https://doi.org/10.1007/s00357-022-09413-z

[37] Alexander Wong, Mahmoud Famuori, Mohammad Javad Shafiee, Francis Li, Brendan Chwyl, and Jonathan Chung. 2019. YOLO Nano: a Highly Compact You Only Look Once Convolutional Neural Network for Object Detection. arXiv (Cornell University) (January 2019). DOI:https://doi.org/10.48550/arxiv.1910.01271

[38] Rikiya Yamashita, Mizuho Nishio, Richard Kinh Gian Do, and Kaori Togashi. 2018. Convolutional Neural networks: an Overview and Application in Radiology. Insights into Imaging 9, 4 (June 2018), 611–629. DOI:https://doi.org/10.1007/s13244-018-0639-9

[39] Muhammad Yaseen. 2024. What is YOLOv8: An In-Depth Exploration of the Internal Features of the Next-Generation Object Detector. arXiv (Cornell University) (August 2024). DOI:https://doi.org/10.48550/arxiv.2408.15857

[40] Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. 2014. How transferable are features in deep neural networks? (November 2014). DOI:https://doi.org/10.48550/arxiv.1411.1792

[41] Zahra Shakoori, Ahmadreza Mehrabian, Dariush Minai, Farid Salmanpour, and Farzaneh Khajoei Nasab. 2023. Assessing the quality of bee honey on the basis of melissopalynology as well as chemical analysis. PLOS ONE 18, 8 (August 2023), e0289702–e0289702. DOI:https://doi.org/10.1371/journal.pone.0289702

[42] 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 (October 2023), 100850. DOI:https://doi.org/10.1016/j.fochx.2023.100850

A SUPPLEMENTARY DATA

Table 5: Comparative CNN model results for pollen classification.

Model / Strategy	Key Feature	Test Acc	Test F1	Val Acc
Targeted ConvNeXt + Full TTA	Augmented inference	94.66%	91.18%	–
Targeted ConvNeXt (Tiny)	Label smoothing + classifier	94.43%	91.58%	94.97%
ConvNeXt Base (Advanced)	Larger model + multi-techniques	94.43%	93.20%	92.51%
ConvNeXt Tiny (Baseline)	Standard setup	93.51%	93.61%	92.48%
ConvNeXt Tiny + CrossEntropy	Enhanced training	93.42%	92.91%	94.41%
ConvNeXt Tiny + Label Smooth	Label smoothing	92.61%	92.19%	93.42%
ConvNeXt Small + Focal Loss	Larger model	92.54%	92.12%	91.85%
ConvNeXt Tiny + Focal Loss	Focal loss + sampling	92.42%	91.52%	93.20%
ResNet50 + Focal Loss	Alternative architecture	92.11%	90.90%	90.15%
Enhanced ConvNeXt Tiny	Over-engineered pipeline	91.96%	91.18%	88.28%
ConvNeXt Tiny + Focal (v2)	Repeat focal loss	91.64%	90.43%	91.61%
EfficientNet-B3 + Focal Loss	Advanced architecture	89.04%	88.26%	89.73%

B ADDITIONAL FIGURES

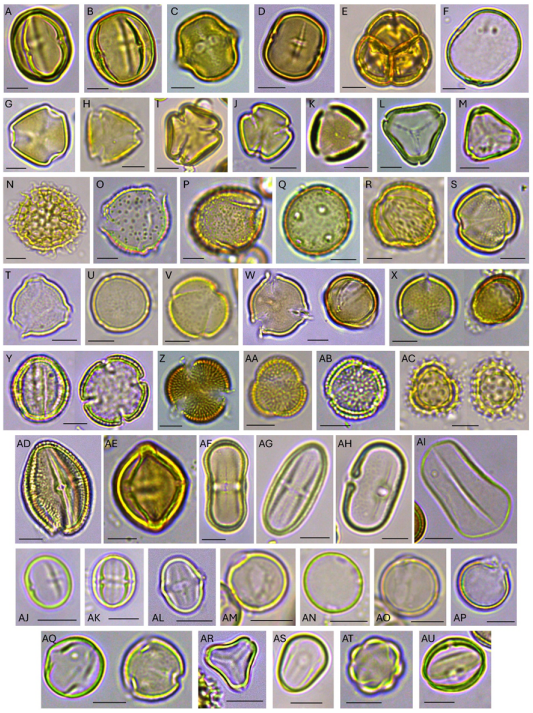

Figure 10: Some from the 77 distinct pollen types show the morphological complexity handled by the POL-ID system.

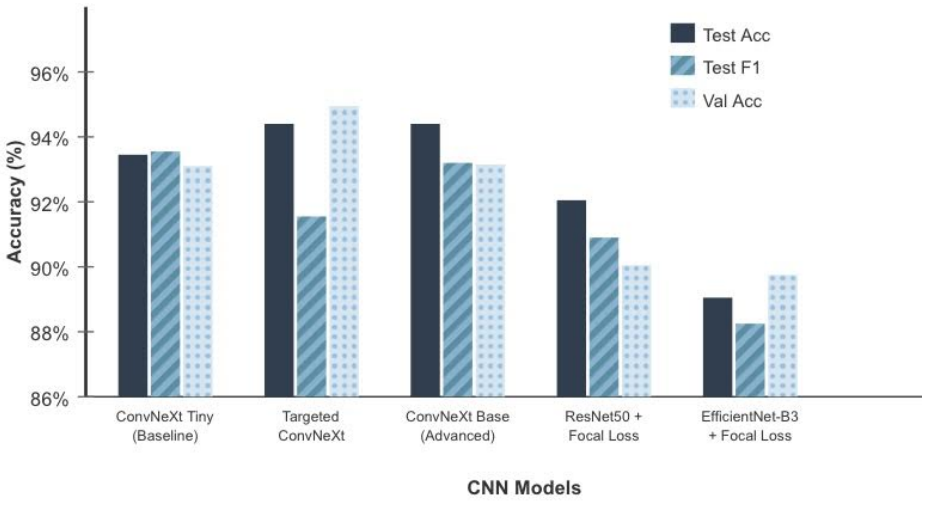

Figure 11: Comparison of different CNN architectures.