Kafkas Universitesi Veteriner Fakultesi Dergisi
Journal Home-Page: http://vetdergikafkas.org
E-ISSN: 1309-2251

Kafkas Univ Vet Fak Derg
31 (5): 653-659, 2025
DOI: 10.9775/kvfd.2025.34544

RESEARCH ARTICLE

Comparative Performance of Convolutional Neural Network Models in Wing Morphometric Classification of Honey Bee Populations Across Europe

Berkant İsmail YILDIZ [1] (*)

[1] Akdeniz University, Faculty of Agriculture, Department of Agricultural Biotechnology, TR-07058 Antalya - TÜRKİYE

(*) **Corresponding author:**
Berkant İsmail Yıldız
Cellular phone: +90 542 516 29 85
E-mail: berkantyildizz@gmail.com

How to cite this article?

Yıldız Bİ: Comparative performance of convolutional neural network models in wing morphometric classification of honey bee populations across Europe. *Kafkas Univ Vet Fak Derg,* 31 (5): 653-659, 2025.
DOI: 10.9775/kvfd.2025.34544

Article ID: KVFD-2025-34544
Received: 25.05.2025
Accepted: 06.10.2025

Abstract

Accurate identification of *Apis mellifera* populations is essential for conserving biodiversity, optimizing breeding programs, and understanding environmental adaptation processes. Traditional morphometric approaches, while informative, face significant limitations due to their labor-intensive nature and inefficiency when applied to large-scale datasets. To address these challenges, this study applies a comparative evaluation of Convolutional Neural Network (CNN) models for country-level classification of honey bee populations using forewing images. A total of 2.500 high-resolution forewing images-500 from each of Croatia, Poland, Romania, Spain, and Greece-were selected to represent diverse geographical regions. Following biologically appropriate preprocessing and data augmentation, the images were analyzed through a comparative evaluation of three pre-trained CNN models - VGG16, InceptionV3, and ResNet50. All models were fine-tuned through transfer learning, and classification performance was systematically assessed using accuracy, precision, recall, and F1-score metrics. Among the evaluated models, VGG16 achieved the highest classification accuracy at 95%, outperforming InceptionV3 and ResNet50. These results highlight not only the high predictive power of CNN models in morphologically distinguishing honey bee populations, but also demonstrate the relative strengths of different CNN models. Furthermore, the study underscores the advantages of automated CNN-based workflows over manual morphometric methods in terms of speed, objectivity, and scalability. By integrating CNN models into morphometric analysis, this research provides robust and reproducible tools for apicultural studies, population-level biodiversity monitoring, and applied breeding strategies.

Keywords: Automated population classification, Convolutional Neural Networks (CNNs), Deep learning in biodiversity, Honey bee, Wing morphometrics

INTRODUCTION

The western honey bee *(Apis mellifera)* plays a central role in sustaining pollination networks that underpin the functioning of natural ecosystems, while also serving as an indispensable component of global agricultural production systems. Pollination services directly influence approximately 75% of flowering plant species and contribute to around 35% of global food production, thereby supporting both biodiversity and food security worldwide [1]. However, in recent decades, various stressors-including climate change, pesticide use in agriculture, and habitat fragmentation-have led to significant declines in local honey bee populations. These threats pose serious risks not only at ecological but also economic and genetic levels, challenging the long-term sustainability of apiculture and pollination services [2,3]. In this context, accurate identification and monitoring of local genetic lineages are crucial for understanding intra-species adaptation dynamics and for developing strategic conservation efforts to safeguard native populations. Traditional morphometric analyses have long been employed as the primary approach for identifying subspecies and geographic variants of honey bees. Among these, forewing venation patterns have historically served as key morphological indicators for differentiating populations [4]. Nevertheless, this method is increasingly inadequate for modern biometric applications due to its dependency on expert interpretation, labor-intensive and time-consuming processes, and limited scalability for large datasets. Furthermore, the reliance on visual assessments introduces subjectivity, which in turn compromises the reproducibility and standardization of results [5].

In recent years, rapid advancements in artificial intelligence and computer vision have enabled the fast, objective, and high-throughput analysis of morphological data. These developments form the foundation for automated systems that increasingly complement human expertise in image-based classification tasks. Convolutional Neural Networks (CNNs) have emerged as powerful and flexible models that overcome the limitations of traditional morphometric approaches, owing to their data-driven learning strategies and multi-layered architectures [6]. CNNs are implemented through diverse architectures that emphasize different strategies for feature extraction and network depth. Prominent examples used in transfer learning include VGG16, which emphasizes a simple yet deep layer organization; ResNet50, which introduced residual connections to allow the training of very deep networks without degradation problems; and InceptionV3, which incorporates multi-scale convolutional filters for efficient feature extraction [7]. A key strength of CNNs lies in their ability to autonomously learn both local and global patterns embedded in image data, eliminating the need for manual feature engineering. Particularly in symmetrical structures such as wing morphology, CNNs effectively capture subtle variations-even those imperceptible to the human eye-that carry taxonomic significance, thus demonstrating high discriminative power in complex morphological pattern recognition tasks [8-10]. These capabilities not only enhance classification performance but also meet core requirements of modern biometric analysis, such as reproducibility, scalability, and adaptability. Consequently, CNN-based systems signify a paradigm shift in morphometric evaluation, paving the way for more comprehensive and objective strategies with the potential to progressively replace traditional methods [11].

Various CNN-based approaches have been proposed in the literature for the classification of honey bee subspecies. For example, De Nart et al.[12] compared different CNN models using 9.887 wing images from seven subspecies and reported that ResNet50 achieved an accuracy above 94%. Similarly, the DeepWings© system developed by Rodrigues et al.[13] employed a hybrid CNN–SVM framework on images from 26 subspecies and achieved an average accuracy of 86.6%. Oleksa et al.[14] on the other hand, made a significant contribution to addressing the issue of data sharing in this field by providing a large-scale open-access dataset comprising 26.481 honey bee wing images collected from different regions across Europe. However, this open-access dataset has so far been utilized mainly for classical morphometric or statistical analyses, and applications of multi-class CNN-based classification remain absent. More broadly, most existing studies have concentrated on subspecies-level identification, leaving the automated classification of geographic population variation largely underexplored.

In this study, a total of 2.500 forewing images of honey bees from Croatia, Poland, Romania, Spain, and Greece were used to achieve automated morphological classification of populations at the country level. To this end, three different CNN architectures-VGG16, InceptionV3, and ResNet50-were implemented through transfer learning, and the classification performance of each model was evaluated based on metrics such as accuracy, precision, recall, and F1-score. The study aims to fill a critical gap in the literature by focusing on geographic population-level classification, which has received limited attention, and by providing a comparative performance analysis of CNN-based models. This approach offers the potential to overcome the limitations of traditional morphometric methods by enabling faster, more objective, and scalable classification processes.

Material and Methods

Ethical Statement

This study did not involve any procedures requiring ethical approval.

Material

In this study, a total of 2.500 forewing images of worker honey bees *(Apis mellifera)* from various regions of Europe were used. The images were obtained from a public domain dataset published by Oleksa et al. and hosted on the Zenodo platform [14]. The original dataset contains 26.481 forewing images collected from 13 European countries; however, for the purposes of this study, only the data from five countries-Croatia, Poland, Romania, Spain, and Greece-with sufficient sample sizes were included in the analysis *(Fig. 1)*.

Fig 1. Locations from which honey bee samples were collected

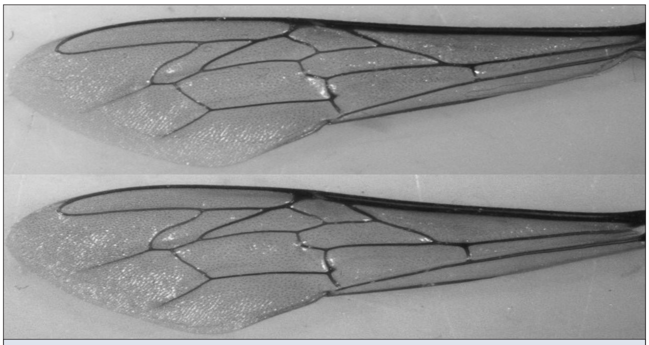

Fig 2. Example of a honey bee forewing with the image quality used in the analyses

For each country, a balanced subset consisting of 500 wing images was created by applying a stratified random sampling procedure across multiple geographic locations included in the dataset. This approach ensured that each country class was equally represented while partially reflecting its internal geographic variation. The images were obtained by mounting the wings of worker bees between microscopic slides and capturing them digitally under controlled optical conditions. All images were in high-resolution. PNG format and were used directly for analysis without any manual annotation or segmentation procedures *(Fig. 2)*. The image files followed a naming convention beginning with ISO 3166-1 alpha-2 country codes (e.g., GR-0001-wing01.png) and were organized into separate. ZIP archives for each country.

Preprocessing

A total of 2.500 wing images in .PNG format were subjected to preprocessing to standardize their dimensions prior to analysis. All images were resized to an input resolution of 224 × 224 pixels and converted into three-channel tensors in RGB format. Pixel values were normalized from the original 0-255 range to a 0-1 scale. No manual segmentation, background removal, or landmark annotation was applied; the images were analyzed in their raw form.

Data Augmentation

To address the limited number of samples in the dataset and to better capture intra-country morphological variation, data augmentation techniques were applied. This process aimed to reduce the risk of overfitting during model training and to enhance the model's ability to recognize generalizable morphological patterns.

From each original wing image, synthetic variations were generated using the following transformations:

- Random horizontal flipping
- Random rotation up to ±15°
- Random width and height shifts of up to ±10%
- Random zooming within a ±10% range
- Minor pixel brightness variations (brightness range: 0.8-1.2)

These transformations were implemented in real time (on-the-fly) on the training data using the ImageDataGenerator class from the Keras library. Data augmentation was applied only to the training set; the validation and test sets were evaluated in their raw, unaltered forms.

CNN Architectures

In this study, three widely used Convolutional Neural Network (CNN) architectures were evaluated for the classification of honey bee wing images: VGG16, ResNet50, and InceptionV3. These models represent different design principles: VGG16 with its simple but deep layer organization is a common reference in transfer learning; ResNet50 introduces residual connections that allow the training of very deep networks without degradation problems; and InceptionV3 employs multi-scale convolutional filters within its modules, enabling efficient extraction of diverse features from complex images [7]. All architectures were implemented through transfer learning using pre-trained ImageNet weights. During training, the initial layers were frozen and only the final layers were fine-tuned for task-specific adaptation.

The input dimensions were standardized to 224 × 224 × 3 across all models. The original classification layers were replaced with a fully connected dense layer of five neurons, each representing one country class, with softmax activation applied to generate probabilities for class membership.

For training, the Adam optimization algorithm was employed with a learning rate of 0.0001. Each architecture was trained independently, and classification performance was evaluated separately to ensure fair comparison.

Model Training

The dataset was divided into 80% training and 20% test sets, with 10% of the training data reserved for validation. Stratified sampling was applied to preserve class distribution across subsets. All CNN models were trained independently for a maximum of 50 epochs with a batch size of 32. To prevent overfitting, early stopping was applied, terminating training if validation loss (val_loss) did not improve for five consecutive epochs. The Adam optimizer was used with a learning rate of 0.0001, and categorical cross-entropy was selected as the loss function due to the multi-class classification task and its compatibility with the softmax activation function. The ModelCheckpoint callback was employed to save the weights corresponding to the epoch with the best validation performance.

All analyses were conducted in Python on the Google Colab platform. Models were implemented with TensorFlow 2.11 and Keras, supported by NumPy, Scikit-learn, Matplotlib, and Seaborn libraries, and training was performed on an NVIDIA Tesla T4 GPU-enabled system.

Evaluation Metrics

To evaluate the classification performance of the models, several metrics were computed on the test dataset. These included accuracy, precision, recall, and F1-score. To provide a comprehensive assessment of each model's ability to correctly identify all classes, these metrics were reported both per class and as macro-averaged scores (*Fig. 3*).

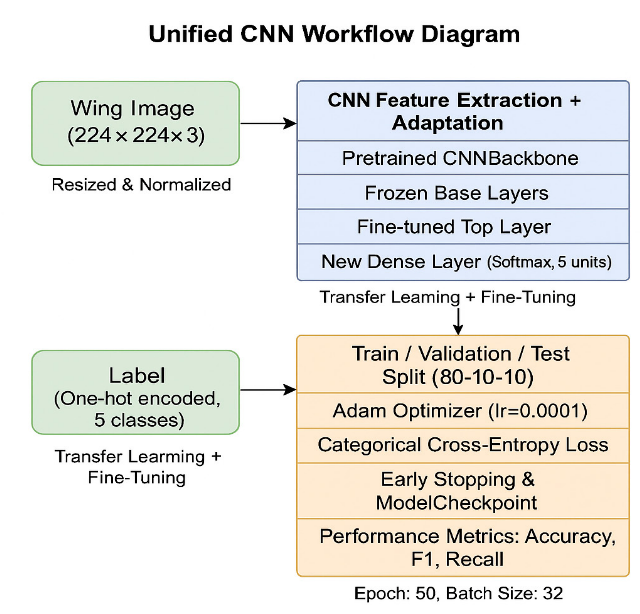

Fig 3. Workflow pipeline used for wing image classification using VGG16, ResNet50, and InceptionV3 architectures

In addition, confusion matrices and Receiver Operating Characteristic (ROC) curves were visualized for each model. Class-wise recall and specificity levels were analyzed in detail. The confusion matrices were used to identify which classes were most frequently misclassified with one another, while the ROC curves and the corresponding Area Under the Curve (AUC) values reflected the overall discriminative power of the models.

PCA-Based Morphological Distribution Analysis

In addition to the classification performance achieved by the deep learning models, Principal Component Analysis (PCA) was performed to further explore inter-individual morphological patterns. For this analysis, shape features derived from each wing image were computed using Hu moments. Hu moments consist of seven statistical values that are invariant to geometric transformations such as rotation, scaling, and translation, and are commonly used to describe object shapes in a stable manner. Each individual was represented as a seven-dimensional feature vector based on the computed Hu moments, and these vectors were subsequently reduced to two dimensions using the PCA algorithm. The resulting PCA plots visually illustrated morphometric similarities and differences among individuals and allowed for the analysis of class-level overlaps.

RESULTS

The classification performance of the three Convolutional Neural Network (CNN) architectures is summarized in *Table 1*. Among them, VGG16 achieved the highest accuracy at 95%, with precision, recall, and F1-score values of 0.95, reflecting balanced and consistent recognition across all classes. InceptionV3 followed with 93% accuracy and comparable precision–recall metrics (≈0.93). ResNet50 showed the lowest overall accuracy (90%) but maintained a relatively high precision (0.93), suggesting a more conservative classification approach that may have reduced false positives.

As shown in *Fig. 4-A,B,C*, the confusion matrices provide a comparative overview of class-wise discrimination. VGG16 reached 100% accuracy for Spain (ES) and Greece (GR), with precision and recall values of 1.00, likely due to distinctive morphological traits in these populations. In contrast, classification errors occurred in Croatia (HR), where 14 samples were misclassified as Poland (PL) (recall = 0.86), and in PL, where four samples were assigned to HR (recall = 0.96). Romania (RO) showed moderate confusion with seven misclassifications into HR and PL. For ResNet50 *(Fig. 4-B)*, ES, GR, and RO were perfectly classified, but minor errors occurred in HR (recall = 0.97). The greatest difficulty was observed in PL: only 55 of 100 samples were correctly identified, with 45 misclassified as HR, leading to sharp declines in both precision and recall. The InceptionV3 model *(Fig. 4-C)* also achieved 100% accuracy for ES and GR. However, HR showed weaker performance, with 23 samples classified as PL and one as RO (recall = 0.76). Precision in PL was reduced by nine misclassifications from HR, while recall remained relatively high (91/100 correctly classified). For RO, only two misclassifications into HR were observed, and the overall recognition was strong.

Table 1. *Average classification performance metrics by model*

Model	Accuracy	Precision	Recall	F1-Score	Support
VGG16	0.95	0.95	0.95	0.95	500
ResNet50	0.90	0.93	0.90	0.90	500
InceptionV3	0.93	0.93	0.93	0.93	500

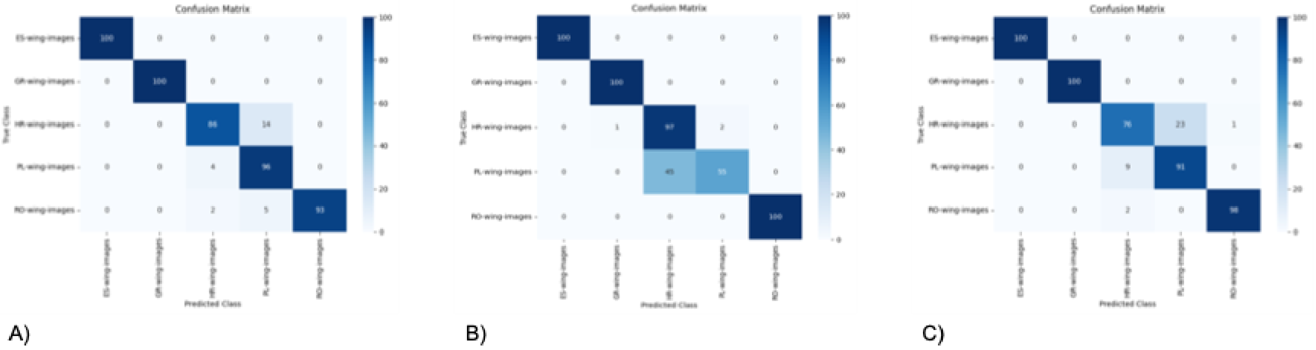

Fig 4. Confusion matrices illustrating the classification performance of different CNN models. (**A**) VGG16, (**B**) ResNet50, and (**C**) InceptionV3. Each matrix displays the actual versus predicted class distributions, highlighting the ability of each model to distinguish among honey bee wing images from different populations

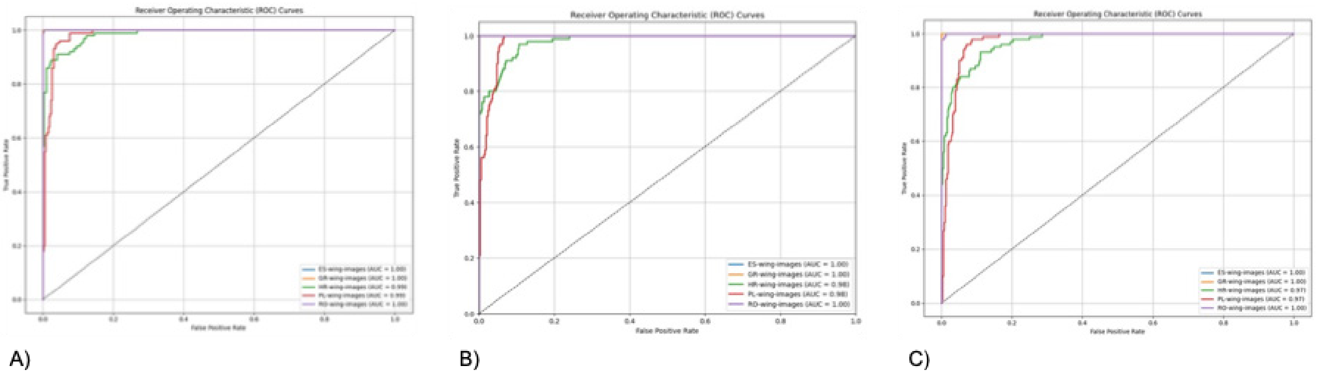

Fig 5. ROC curves demonstrating the classification performance of the CNN models. **(A)** VGG16, **(B)** ResNet50, and **(C)** InceptionV3. The curves illustrate the trade-off between true positive and false positive rates across different classification thresholds. AUC values are reported to quantify the overall discriminative ability of each model

Receiver Operating Characteristic (ROC) curves for the three CNN architectures *(Fig. 5-A,B,C)* demonstrated consistently high inter-class discriminative performance. For Spain (ES), Greece (GR), and Romania (RO), all models achieved an Area Under the Curve (AUC) value of 1.00, confirming perfect separation without classification errors. This finding highlights the distinctiveness of the wing patterns in these populations and the strong ability of CNNs to capture such features. By contrast, slight reductions in AUC values were observed for Croatia (HR) and Poland (PL), with VGG16 achieving the highest scores (0.99), followed by ResNet50 (0.98) and InceptionV3 (0.97). These differences indicate that morphological similarities between HR and PL created a greater challenge for accurate separation. The ROC curve slopes further suggest that InceptionV3 performed comparatively weaker for these classes. Overall, ROC analysis confirmed VGG16 as the most consistent model, ResNet50 as balanced, and InceptionV3 as relatively limited in discriminative ability for certain populations.

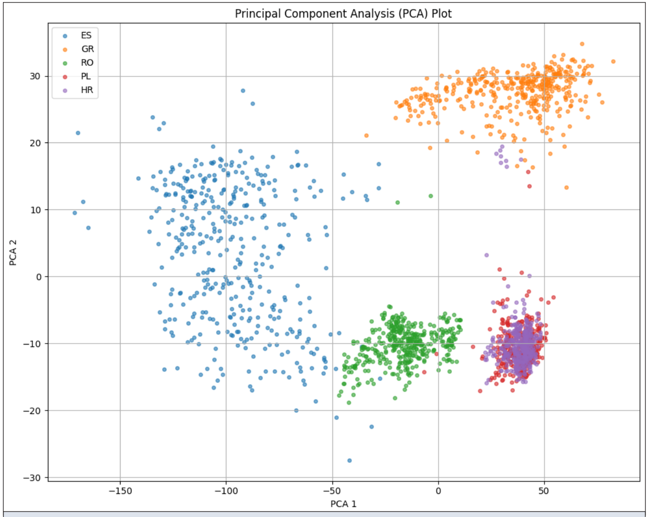

Fig 6. Two-dimensional PCA plot illustrating the morphological variation of honey bee wing images from five different populations. Each point represents a single individual, and colors indicate country-specific groupings (ES, GR, RO, PL, HR). The first two principal components collectively explain 80.09% of the total variance

In addition to the classification analyses, a Principal Component Analysis (PCA) was performed using shape features derived from Hu moments to further explore morphological similarities and differences among the samples. The first two principal components together explained 80.09% of the total variance, indicating that most of the morphological variation was effectively captured.

As shown in *Fig. 6*, the five country groups generally formed distinct clusters; however, a marked overlap was observed between Croatia (HR) and Poland (PL). This overlap was also reflected in the CNN misclassification patterns, where up to 45% of PL samples were assigned to HR by ResNet50.

Discussion

The findings of this study demonstrated that CNN-based models can classify honey bee populations with high overall accuracy, particularly when using wing image data. Among the evaluated architectures, VGG16 consistently outperformed ResNet50 and InceptionV3 across accuracy, precision, recall, and F1-score metrics, indicating its robustness and balance in handling class distinctions. These results align with previous studies, such as De Nart et al.[12], which demonstrated the applicability of CNN-based classification to honey bee wing images, supporting the notion that such models offer higher-resolution discrimination than traditional morphometric approaches. Taken together, these findings indicate that while all models produced highly accurate results for certain classes, classification performance declined notably due to recurring misclassifications between the HR and PL classes. The consistent confusion between Croatia (HR) and Poland (PL) may be rooted not only in algorithmic limitations but also in biological similarity. According to Mahalanobis distance values reported by Oleksa et al.[14], the distance between the HR and PL populations was calculated as 3.91-the smallest among the five countries analyzed in this study. Thus, it can be inferred that these classification errors reflect not only model-specific weaknesses but also genuine morphometric proximity, which naturally influences the learning process of the models.

Previous studies have employed a variety of CNN architectures for classifying bee wing images, thereby providing useful comparative benchmarks. De Nart et al.[12] evaluated ResNet50, MobileNetV2, InceptionV3, and Inception-ResNetV2 architectures on approximately 9,887 honey bee wing images and reported accuracies exceeding 92%, with Inception-based models achieving the strongest performance in many classes. Similarly, Spiesman et al.[15] demonstrated that convolutional neural networks, specifically EfficientNetV2L, can achieve high accuracy (up to 98.1%) when classifying challenging bee taxa using forewing images, reinforcing the value of morphological image-based approaches for species-level identification. In a broader insect context, Sauer et al.[16] developed a CNN for distinguishing mosquito species based solely on wing images and achieved a macro-F1 score of about 0.90 using RGB images, demonstrating that even highly similar wing venation patterns can be reliably differentiated with well-trained architectures. In the present study, VGG16 delivered the best performance. This outcome may be attributed to the model's balance between complexity and regularization, which allowed it to generalize effectively on a limited dataset without overfitting. ResNet50, while robust for deep feature learning due to its residual connections, tended to perform more conservatively in morphologically similar classes, leading to reduced recall. InceptionV3, although designed to capture multi-scale features, was less effective than VGG16 in separating the HR-PL populations. Collectively, these findings suggest that in contexts where repetitive and fine-scale morphological traits dominate-such as forewing venation-simpler architectures may offer an optimal trade-off between capacity and generalization. The observed performance differences support the view that architecture choice should be carefully aligned with the morphological complexity of the problem under investigation.

Beyond the performance comparisons, another important distinction is methodological. Unlike landmark-based approaches such as DeepWings© software developed by Rodrigues et al.[13] and Garcia et al.[17] or similar systems that rely on predefined vein coordinates, our method did not require any manual extraction of morphometric landmarks. Instead, the CNN models directly processed raw wing images, learning discriminative features in a data-driven manner. This represents a methodological shift in wing morphometrics, as it removes observer-dependent steps and allows the models to capture subtle and potentially more informative morphological patterns that may be overlooked in landmark-based analyses. Such an approach enhances objectivity, reduces preprocessing effort, and expands the scalability of morphometric studies to large datasets.

The applications of artificial intelligence in apiculture extend far beyond classification. For example, DeepBee© achieved over 98% accuracy in detecting eggs, larvae, and honey, outperforming traditional observation [18]. Other approaches, such as those of Voudiotis et al.[19], integrated cameras with deep learning to identify Varroa-infested bees with 86% accuracy. Deep learning has also been applied to more complex biological features: Lösel et al.[20] distinguished brain structures of bees and wasps using micro-CT and CNNs. In addition, Kongsilp et al.[21] combined Mask R-CNN with Kalman filtering to track waggle dances within hives. Collectively, these studies highlight the versatility of deep learning in tackling diverse apicultural challenges, ranging from colony health monitoring to behavioral analysis.

This study demonstrated that CNN-based models can classify honey bee populations with high accuracy based on wing morphology, with the VGG16 architecture exhibiting particularly balanced and reliable performance. The recurrent overlaps observed between morphologically similar populations, such as HR and PL, reflect not only model-related challenges but also genuine biological proximities. Beyond these results, the present work introduces a distinctive contribution by demonstrating that accurate classification can be achieved without manual landmark extraction, relying instead on fully automated, data-driven feature learning from raw wing images. This methodological simplification enhances objectivity and scalability, making the approach more suitable for large-scale morphometric applications. Future research may further improve class-level discrimination and generalization capacity by incorporating more diverse datasets and advanced modeling strategies, but the present findings already establish CNN-based workflows as a robust and innovative framework for population-level analyses in apicultural research.

DECLARATION

Availability of Data and Materials: The dataset used in this study is publicly available on Zenodo at https://zenodo.org/records/7244070.

Acknowledgements: The author declares no acknowledgements.

Funding Support: The author received no funding for this study.

Competing Interest: The author declares that there are no competing interest.

Declaration of Generative Artificial Intelligence (AI): The author declare that the article, tables and figures were not written/created by AI and AI-assisted Technologies.

REFERENCES

1. Klein AM, Vaissière BE, Cane JH, Steffan-Dewenter I, Cunningham SA, Kremen C, Tscharntke T: Importance of pollinators in changing landscapes for world crops. *Proc Biol Sci,* 274 (1608): 303-313, 2007. DOI: 10.1098/rspb.2006.3721

2. Potts SG, Imperatriz-Fonseca V, Ngo HT, Aizen MA, Biesmeijer JC, Breeze TD, Dicks LV, Garibaldi LA, Hill R, Settele J, Vanbergen AJ: Safeguarding pollinators and their values to human well-being. *Nature,* 540 (7632): 220-229, 2016. DOI: 10.1038/nature20588

3. Vercelli M, Novelli S, Ferrazzi P, Lentini G, Ferracini C: A qualitative analysis of beekeepers' perceptions and farm management adaptations to the impact of climate change on honey bees. *Insects,* 12 (3):228, 2021. DOI: 10.3390/insects12030228

4. Ruttner F: Biogeography and Taxonomy of Honeybees. Springer Berlin, Heidelberg, 1988.

5. Tofilski A: Using geometric morphometrics and standard morphometry to discriminate three honeybee subspecies. *Apidologie,* 39 (5): 558-563, 2008. DOI: 10.1051/apido:2008037

6. Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, Farhan L: Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *J Big Data,* 8:53, 2021. DOI: 10.1186/s40537-021-00444-8

7. Zhao X, Wang L, Zhang Y, Han X, Deveci M, Parmar M: A review of convolutional neural networks in computer vision. *Artif Intell Rev,* 57 (4):99, 2024. DOI: 10.1007/s10462-024-10721-6

8. Buschbacher K, Ahrens D, Espeland M, Steinhage V: Image-based species identification of wild bees using convolutional neural networks. *Ecol Inform,* 55:101017, 2020. DOI: 10.1016/j.ecoinf.2019.101017

9. Geldenhuys DS, Josias S, Brink W, Makhubele M, Hui C, Landi P, Bingham J, Hargrove J, Hazelbag MC: Deep learning approaches to landmark detection in tsetse wing images. *PLoS Comput Biol,* 19 (6):e1011194, 2023. DOI: 10.1371/journal.pcbi.1011194

10. Ansari MA, Crampton A, Garrard R, Cai B, Attallah M: A Convolutional Neural Network (CNN) classification to identify the presence of pores in powder bed fusion images. *Int J Adv Manuf Technol,* 120 (7): 5133-5150, 2022. DOI: 10.1007/s00170-022-08995-7

11. LeCun Y, Bengio Y, Hinton G: Deep learning. *Nature,* 521 (7553): 436-444, 2015. DOI: 10.1038/nature14539

12. De Nart D, Costa C, Di Prisco G, Carpana E: Image recognition using convolutional neural networks for classification of honey bee subspecies. *Apidologie,* 53 (1):5, 2022. DOI: 10.1007/s13592-022-00918-5

13. Rodrigues PJ, Gomes W, Pinto MA: DeepWings©: Automatic wing geometric morphometrics classification of honey bee *(Apis mellifera)* subspecies using deep learning for detecting landmarks. *Big Data Cogn Comput,* 6 (3):70, 2022. DOI: 10.3390/bdcc6030070

14. Oleksa A, Căuia E, Siceanu A, Puškadija Z, Kovačić M, Pinto MA, Rodrigues PJ, Hatjina F, Charistos L, Bouga M, Prešern J, Kandemir İ, Rašić S, Kusza S, Tofilski A: Honey bee *(Apis mellifera)* wing images: A tool for identification and conservation. *GigaScience,* 12:giad019, 2023. DOI: 10.1093/gigascience/giad019

15. Spiesman BJ, Gratton C, Gratton E, Hines H: Deep learning for identifying bee species from images of wings and pinned specimens. *PloS one*, 19 (5):e0303383, 2024. DOI: 10.1371/journal.pone.0303383

16. Sauer FG, Werny M, Nolte K, Villacañas de Castro C, Becker N, Kiel E, Lühken R: A convolutional neural network to identify mosquito species (Diptera: Culicidae) of the genus Aedes by wing images. *Sci Rep,* 14 (1):3094, 2024. DOI: 10.1038/s41598-024-53631-x

17. García CAY, Rodrigues PJ, Tofilski A, Elen D, McCormak GP, Oleksa A, Henriques D, Ilyasov R, Kartashev A, Bargain C, Fried B, Pinto MA: Using the software DeepWings© to classify honey bees across Europe through wing geometric morphometrics. *Insects,* 13 (12): 1132, 2022. DOI: 10.3390/insects13121132

18. Alves TS, Pinto MA, Ventura P, Neves CJ, Biron DG, Junior AC, De Paula Filho DL, Rodrigues PJ: Automatic detection and classification of honey bee comb cells using deep learning. *Comput Electron Agric,* 170:105244, 2020. DOI: 10.1016/j.compag.2020.105244

19. Voudiotis G, Moraiti A, Kontogiannis S: Deep learning beehive monitoring system for early detection of the Varroa mite. *Signals,* 3 (3): 506-523, 2022. DOI: 10.3390/signals3030030

20. Lösel PD, Monchanin C, Lebrun R, Jayme A, Relle JJ, Devaud JM, Heuveline V, Lihoreau M: Natural variability in bee brain size and symmetry revealed by micro-CT imaging and deep learning. *PLoS Comput Biol,* 19 (10):e1011529, 2023. DOI: 10.1371/journal.pcbi.1011529

21. Kongsilp P, Taetragool U, Duangphakdee O: Individual honey bee tracking in a beehive environment using deep learning and Kalman filter. *Sci Rep,* 14 (1):1061, 2024. DOI: 10.1038/s41598-023-44718-y