



U-Net 기반 날개 시맥의 형태분석을 통한 꿀벌 아종 자동 분류

권순호[†], 조희원^{1,†}, 이철희^{1,*}, 정철의^{*}

국립경국대학교 식물학과, ¹국립경국대학교 컴퓨터공학과

Automatic Classification of Honey Bee Subspecies via U-Net-assisted Morphometrics

Sunho Kwon[†], Huiwon Cho^{1,†}, Cheolhee Lee^{1,*} and Chuleui Jung^{*}

Department of Plant Medicinals, GyeongKuk National University, Andong 36729, Republic of Korea

¹Department of Computer Engineering, GyeongKuk National University, Andong 36729, Republic of Korea

Abstract

Identifying honey-bee subspecies is critical for preserving genetic diversity and understanding behavioral traits related for beekeeping. Conventional morphometrics are labor-intensive and relatively inaccurate, while molecular methods provide high accuracy but at substantial cost. Geometric morphometrics offer a cost-effective alternative but require manual annotation of 19 landmarks. This study proposes an automated pipeline integrating U-Net-based landmark detection, Hungarian matching, GPA-based normalization, and supervised classification. A 210-dimensional feature vector was constructed from aligned coordinates (38D), pairwise distances (171D), and centroid size (1D). Using ~400 forewing images of *Apis mellifera carnica*, *A. m. ligustica*, and *A. m. carpathica*, performance was evaluated via nested cross-validation. Landmark detection achieved Precision 0.844, Recall 0.777, Dice 0.844, and IoU 0.682. Classification accuracy reached 86.4% (LDA), 86.3% (SVM), 77.8% (Random Forest) and 85.2% (MLP). These results demonstrate that automated landmark detection achieved practical accuracy and reduced manual effort, while linear and margin-based classifiers outperformed tree-based models for subspecies classification.

Keywords

Subspecies, Landmarks, U-Net, Geometric morphometrics

서 론

꿀벌은 1만년 이상 인간과 공존하며 꿀, 밀랍, 프로폴리스, 화분, 봉독, 로열젤리를 포함한 여러 봉산물들을 인간에게 제공해 왔다(Ruttner, 2013; Roffet-Salque *et al.*, 2015; Pasupuleti *et al.*, 2017; Jung, 2022). 또한 꿀벌의 화분매개 활동은 전 세계 100대 주요 식량작물의 약 75종의 품질에 직접적인 영향을 미치고, 농업 생산량의 약 30% 이상이 화분매개 활동에 의존하는 것으로 추정된다(Klein *et al.*, 2007; Geldmann and González-Varo, 2018).

국내 농업 생산에서 화분매개의 경제적 가치는 약 6조 원으로 추정된 바 있다(Jung, 2008). 전 세계 2만 종 이상의 벌목 곤충 중 약 50종이 인간에 의해 이용되고 관리되고 있으며, 그 중 특히 꿀벌(*Apis* sp.)이 가장 광범위하게 사용되고 있다(Potts *et al.*, 2016). 꿀벌은 극지방 등 극단적인 환경을 제외한 대부분 지역에 분포하고(Ruttner, 2013), 지역 격리를 통해 다수의 아종으로 분화하였다(Ruttner, 1988; Meixner *et al.*, 2013). 서양꿀벌(*Apis mellifera*)은 유럽과 아프리카

에 자연 분포하였고 현재까지 33아종이 보고되었으며, 지역 기원에 따라 A(아프리카), C(남·동유럽), O(중동 및 중앙아시아), M(서·북유럽) 4개의 주요 계열로 구분되어 왔으며, 최근 연구에 따르면, Y(서아시아), L(이집트), U(마다가스카르)를 포함하여 7개의 계열이 존재하는 것으로 제안되고 있다(Ruttner, 1988; Moritz *et al.*, 2005; Chapman *et al.*, 2016; Ilyasov *et al.*, 2020; Dogantzis *et al.*, 2021).

각 아종은 기후 적응성, 공격성, 채집 행동, 수밀력, 온순성 등에서 뚜렷한 차이를 보이며(Dilday, 2022), 이러한 특성을 바탕으로 아종 간 교배를 통한 선발 육종으로 우수 형질을 가진 품종이 보급·관리되고 있다(Adam *et al.*, 1977; Lee *et al.*, 2015). 그러나 여왕의 다수의 수벌과 공중 교미로 인한 잡종화 우려가 있어 순계 꿀벌 집단의 지속적인 선발 및 관리가 필요하며 이를 위해 신뢰성 있는 아종 분류 체계가 요구된다(Mackensen, 1955; Woyke, 1976; Muñoz and De la Rua, 2021; Akongte *et al.*, 2024).

국내에서는 2005년에서 2007년까지 A, C, D, F, V 계통의 순계 집단 선발이 이루어졌다(Kim *et al.*, 2015; Frunze *et al.*, 2020). 현재 전라북도 부안군 위도면 소재 꿀벌격리육종장에서 순계 꿀벌의 관리가 이루어지고 있으며, 계통별 위생행동, 로열젤리 생산량, 수밀력 등 특성 연구와 선발 육종 노력이 이어지고 있다(Lee *et al.*, 2014; Kim *et al.*, 2015, 2021; Lee *et al.*, 2017; Lee *et al.*, 2019; Kang *et al.*, 2025).

꿀벌의 아종 분류는 고전적인 형태분석법, 기하학적 형태분석법, 분자유전학적 방법 등에 기반하고 있다(Tofilski, 2008; Kandemir *et al.*, 2011; Frunze *et al.*, 2020). 고전적인 형태분석법은 머리 너비, 몸통 길이, 날개 길이 등 신체 부위에 해당하는 36가지 특성을 측정 후 다변량 통계분석을 통해 아종을 식별하는 방법이다(Ruttner, 1988). 그러나 이러한 방법은 측정 항목이 많고 수작업에 의존하여 노동 집약적이라는 한계를 가진다. 기하학적 형태분석법은 꿀벌 앞날개의 시맥 교차점을 사용하여, 획득한 좌표의 편차를 바탕으로 다변량 통계분석(Multi-variate statistical analysis)을 통해 아종을 분류하는 방법이다(Miguel *et al.*, 2011; Ilyasov *et al.*, 2020). 이 과정은 tpsDig, DrawWing, ABIS, IdentiFly 등 별도의 소프트웨어를 사용해 자동 또는 수동으로 수행한다. 이러한 방법은 정확도와 비용 측면에서 장점을 가지지만, 여전히 시맥 교차점 지정 과정에서

사용자의 개입이 필요하다는 한계가 있다.

분자유전학적 방법은 초위성체(microsatellite) 마커와 단일염기다형성(single nucleotide polymorphism, SNP)과 같은 다수의 분자마커를 활용하여, PCR 기반 분석이나 미토콘드리아 DNA(mtDNA) 분석을 통해 유전적 변이와 계통 구조를 탐지하는 접근법이다(Whitfield *et al.*, 2006; Kim *et al.*, 2017). 이러한 방법은 높은 정확도를 제공하지만 비용과 분석 인프라 측면에서 일상적인 활용에는 제약이 있다.

본 연구에서는 수작업 시맥 교차점 지정의 한계를 보완하여 꿀벌 앞날개 이미지로부터 아종을 자동 분류하는 모델을 제시하는 것을 목표로 한다. 이를 위해 (1) 꿀벌 앞날개 사진을 이용해 픽셀 단위 분할에 적합한 U-Net에 학습하고 (2) 학습 모델 추론을 통해 19개의 시맥 교차점 좌표를 자동 산출하며, (3) 산출된 좌표를 210차원의 특징으로 변환 후 입력 특징으로 사용하여 LDA, SVM, Random Forest, MLP 분류기의 성능을 비교한 후 (4) 각 집단별 서열화 분석(ordination)을 통해 자동 분류기의 성능을 검증하였다. 시맥 교차점 탐지를 위한 U-Net 학습에는 아종 구분 없이 구성된 기본 데이터셋을 사용하였고, 분할 성능을 평가하여 시맥 교차점 검출의 성능을 검증하였다. 아종 분류를 위한 분류기 학습에는 *A. m. carnica*, *A. m. ligustica*, *A. m. carpathica*의 국내 유지 계통(A, C, D, F, V)의 앞날개 이미지 약 400장을 활용하였다. 연구의 초점은 개별 계통 간 차이 자체보다는 형태형질에 기반한 아종 간 분류 가능성을 평가하는 데 두었다. 성능 평가는 중첩 교차검증(nested cross-validation)으로 정확도를 기반으로 분류기 간 성능을 비교하였다. 이를 통해 형태 기반 아종 분류의 정확성과 효율성을 동시에 향상시킬 수 있는 자동화된 접근법을 제시하고자 하였다.

재료 및 방법

1. 꿀벌 아종 시료 확보

본 연구에서는 시맥 교차점 탐지를 위해 기존에 촬영하여 보관 중이던 꿀벌 560마리의 앞날개 이미지를 사용하였다(데이터세트 1). 또한 아종 분류를 위해 국내 꿀벌격리육종장(농촌진흥청 국립농업과학원 농업생물부 양봉과 꿀벌위도격리육종장, N 35.591011, E 126.279422)에서 관

Table 1. The information of honeybee subspecies and breeding lines of *A. mellifera* L. used in this study

| Line | Subspecies | n | Color | Import source |
|------|-------------------|----|--------------|-----------------|
| A | <i>ligustica</i> | 68 | Dark yellow | Australia |
| C | <i>ligustica</i> | 77 | Light yellow | Northeast China |
| D | <i>carpathica</i> | 99 | Gray | Northeast China |
| F | <i>ligustica</i> | 70 | Yellow-brown | USA |
| V | <i>carnica</i> | 70 | Dark gray | Europe |

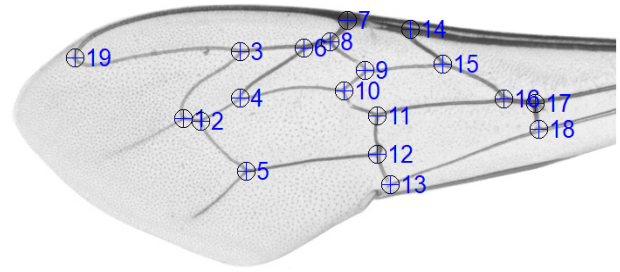
리하는 꿀벌 A, C, D, F, V 총 5계통을 대상으로, 2025년 5월에 각 계통별 3개 봉군에서 15~30마리씩 채집하여 총 384마리를 확보하였다(Table 1, 데이터세트 2). 채집한 순계 꿀벌 샘플은 모두 80% 에탄올에 액침하여 상온 보관하였다. 이후 촬영 시 시료를 건조한 후 핀셋을 이용해 앞날개를 분리하였으며, 실체현미경(KS-208, KOREA LAB TECH, 대한민국) 아래에 고정하여 10배 배율로 이미지를 촬영하였다.

2. 시맥 교차점 기계 인식 학습(자동 분류)

수집된 앞날개 이미지는 연구 목적에 따라 두 개의 데이터세트로 구분하였다. 데이터세트 1은 꿀벌 560마리의 우측 앞날개 이미지를 대상으로 아종을 구분하지 않았다. 각 영상에 Fig. 1에 제시된 앞날개 시맥 교차점(vein junction) 19개 지점을 날개 교차점으로 정의하고 정답 주석을 부여하여 U-Net의 학습 및 성능 평가에 사용하였다. 데이터세트 2는 학습된 U-Net을 이용한 시맥 교차점 추론, GPA 기반 특징 추출, 지도학습 분류기의 학습 및 평가에 사용하였다. 기계 인식 학습(자동 분류) 모델의 경우 데이터세트 2에서 각 순계별로 샘플링된 꿀벌의 60%는 모델 학습에, 나머지 40%는 검증용 서열화 분석에 사용하였다.

3. 시맥 형태 분포를 통한 아종 동정(수동 분류)

수동 분류는 IdentiFly 프로그램(버전 1.8.0)을 사용하여 Fig. 1과 같이 앞날개 시맥의 19개 시맥 교차점 좌표를 지정하였다. IdentiFly는 앞날개 시맥 패턴의 기하학적 형태 계측 지표인 19개의 시맥 교차점을 기반으로, 꿀벌을 4개의 계열(A, C, M, O)과 20개의 아종 수준에서 식별할 수 있는 소프트웨어이다. 분류 분석은 개체 단위가 아닌 군집 단위의 평균 시맥 교차점 좌표를 기반으로 수행하였다(Nawrocka *et al.*, 2018). 이때 사용된 꿀벌 개체는 자동 분

**Fig. 1.** Nineteen landmarks on the forewing of *Apis mellifera*, shown in the IdentiFly program (ver. 1.8.0).

류 모델 학습에 사용된 개체와 동일하다. 수동 랜드마크 지정은 관찰자 간 편차를 최소화하기 위해 연구자 1인이 수행하였으며, IdentiFly에서 제시하는 각 시맥 교차점의 기준 위치를 참고하여 모든 개체에 동일한 기준으로 좌표를 지정하였다. 자동 분류는 개체 단위 앞날개 이미지를 이용하여 시맥 교차점 좌표를 자동 산출한 후 오디네이션 분석을 수행하였으며, 수동 분류는 동일 개체에서 랜드마크를 직접 지정하여 얻은 좌표를 이용해 서열분석을 수행하였다. 최종 성능 비교는 두 방법 모두 서열분석 결과를 바탕으로 집단 단위 평균 정확도와 여분 값을 동일한 방식으로 계산하여 수행하였다.

4. 시맥 교차점 자동 탐지를 위한 시각 인식 기법

기하학적 형태분석법의 특성상 픽셀 수준의 정밀 좌표가 요구된다. 경계 상자(Bounding Box) 기반 객체 검출보다 픽셀 단위로 정확하게 시맥 교차점을 추출할 수 있는 이미지 분할(Image Segmentation) 기법을 선정하였다. 분할 모델은 U-Net (Ronneberger *et al.*, 2015)으로 구성하였다. Fig. 2는 U-Net의 세부 구조를 나타낸 것이다. U-Net은 인코더, Bottleneck, 디코더의 3가지 구조로 이루어져 있다. 인코더는 3×3 합성곱 신경망과 최대 풀링층으로 해상도를 단계적으로 줄여 점차 추상적인 특징을 추출한다. Bottleneck은 네트워크 중간에 위치하여 가장 압축된 특징 맵이며, 높은 수준의 추상적인 특징을 가지고 있다. 디코더는 업샘플링 레이어와 역 합성곱 층을 통해 해상도를 다시 복원한다. 각 층에서 인코더의 특징 맵은 디코더에 스킵 연결(skip connection)로 전달되어 추상 정보를 보존한다.

5. 템플릿 정렬 및 헝가리안 알고리즘

학습된 U-Net은 19개의 시맥 교차점을 자동 추출하지

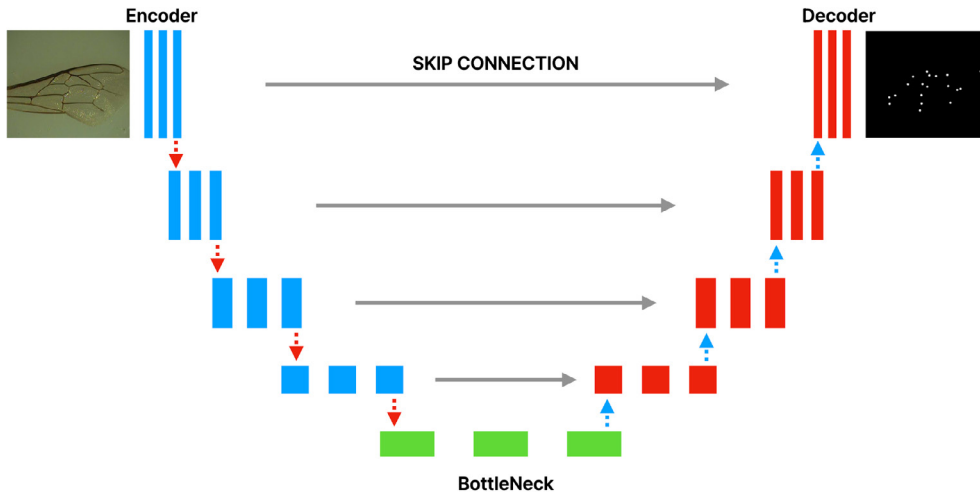


Fig. 2. U-Net architecture for landmark segmentation.

만, 샘플 간 번호 순서의 일관성을 보장하지 않는다. 이후의 아종 분류를 위해서 샘플 간 시맥 교차점의 일관된 번호 정렬이 필요하다. 본 연구에서는 Francoy *et al.* (2008) 및 Fig. 1에서 정의한 19개의 점 템플릿을 사용하여 인덱스를 부여하였다. 이를 위해서 모든 샘플 시맥 분지점의 평균 좌표를 구해 평균 형상을 구한 후 다음 두 단계를 거쳐 순서가 없는 각 샘플의 시맥 교차점 집합을 템플릿에 정합하였다.

1) 끝점 정렬을 이용한 거친 유사변환(coarse alignment)

샘플에서 추출된 19개의 시맥 교차점 중 가장 긴 점쌍을 (1, 19)번의 후보로 지정 후, 만약 템플릿의 (1, 19)번 시맥 교차점 사이의 길이와의 오차가 임계치를 초과하는 경우 주성분축의 양 끝점을 기준으로 재지정하였다. 지정된 두 점을 템플릿의 (1, 19)번에 대응하여 평행이동, 스케일, 회전을 계산 후, 샘플 좌표를 템플릿에 대략적으로 맞춘다.

2) 헝가리안 매칭을 통한 인덱싱

1번과 19번 시맥 교차점을 찾은 후 고정한 상태에서 나머지 17개의 시맥 교차점과 템플릿 사이에서 유클리드 거리 기반 비용 행렬을 계산 후, 헝가리안 알고리즘으로 비용을 최소화하는 일대일 대응을 계산한다. 이 매칭 결과에 따라 샘플 좌표를 재정렬한다. 이로써 모든 샘플은 일관된 인덱스를 가지게 된다.

6. 형태 정규화 및 특징 구성

U-Net에서 추출된 순수 시맥 교차점 좌표는 스케일, 회

전, 위치에 민감하다. 불변성을 가지는 특징을 추출하기 위해서 Generalized Procrustes Analysis (GPA)를 적용하여 순수 형태만 남기도록 정렬하였다. GPA는 Fig. 3과 같이 각 샘플을 공통의 중심으로 평행 이동한 뒤, 센트로이드 크기(Centroid size)가 1이 되도록 스케일을 정규화한 다음, 회전을 최적화하여 공통 공간에 정렬하는 기하학적 형태측정 기법이다. GPA를 적용한 좌표는 중심이 원점으로 이동되고, 스케일이 제거되어, 회전에 불변한 형태 좌표가 된다. 최종 특징 벡터의 차원은 GPA 정렬 좌표를 벡터화한 형태 좌표(38차원)에 더해서, 정렬 좌표에서 모든 점쌍의 유클리드 거리를 계산한 모든 점쌍 간 거리(pairwise distance, 171차원)와 중심점(Centroid)으로부터 얼마나 퍼져 있는지 나타내는 값인 Centroid size (1차원)를 결합하여 총 210차원의 특징으로 구성하였다.

7. 분류기 구성

실험에 사용된 지도학습 분류기는 LDA, SVM, Random Forest, MLP를 사용하였다. LDA는 데이터가 정규분포를 따르고 클래스별 공분산이 동일하다는 가정하에 작동하며, 저차원 공간에서 분류 경계를 단순화한다. SVM은 데이터의 클래스를 구분하는 최적의 초평면을 찾으며, 마진을 최대화하며 일반화 성능을 높이고, 커널 함수를 통해 비선형 데이터도 고차원 공간에서 선형적으로 분리 가능하도록 한다. Random Forest는 여러 개의 결정 트리를 생성 후, 예측 결과를 투표 또는 평균하여 최종 결과를 도출하는 배경 기반 앙상블 학습 알고리즘이다. MLP는 입력층

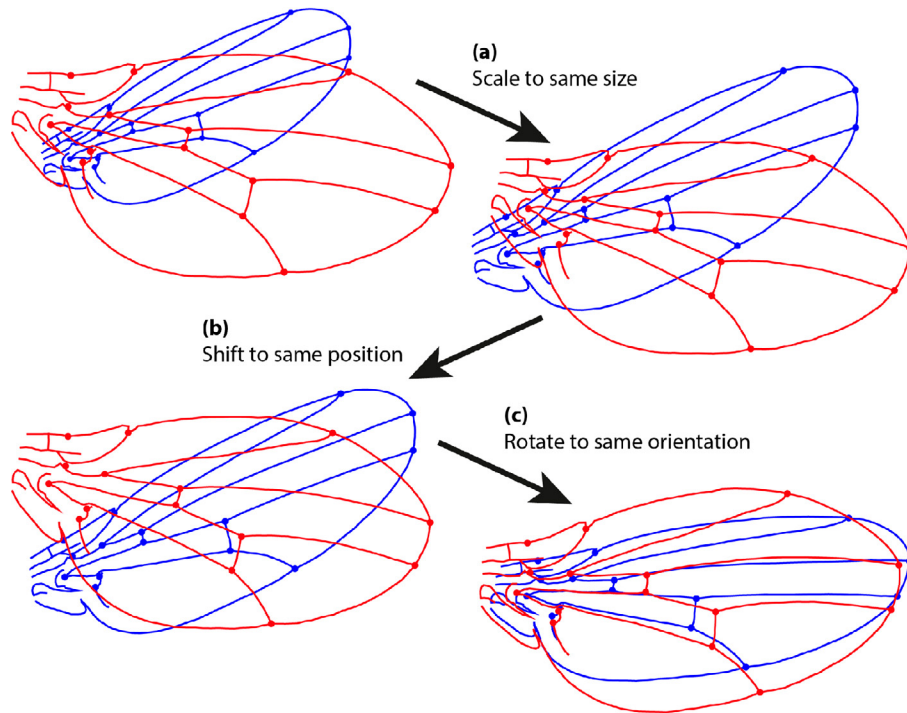


Fig. 3. Generalized procrustes analysis for the morphological standardization of wing venation.

과 출력층 사이에서 하나 이상의 은닉층을 가지는 전방향 신경망으로, 역전파 알고리즘으로 가중치를 학습하여 비선형 관계를 효과적으로 모델링한다. 모든 분류기는 동일한 특징을 입력받아 학습되었다.

8. 실험 설계 및 평가 방법

첫 번째 실험은 준비된 앞날개 이미지인 데이터셋 1을 사용해 U-Net이 19개 시맥 교차점을 정확히 추출할 수 있는지 평가하기 위해 설계되었다. U-Net 모델을 학습하면 가중치(weight) 파일이 여러 개 생성된다. 그중 검증 성능이 가장 좋은 가중치 파일(best.pth)을 최종 추론용으로 선정하였다. 선정된 가중치를 이용해 입력 이미지로부터 19개 시맥 교차점을 추출하였다. 두 번째 실험은 실험 1에서 선정된 best.pth를 고정된 뒤, 데이터셋 2에 대해 추론을 수행해 19개 시맥 교차점 좌표를 얻었다. 이후 템플릿 정렬 및 GPA 정렬과 특징 추출 절차를 거쳐 순수 좌표를 210차원 특징으로 변환하였다. 추론된 210차원 특징을 LDA, SVM, Random Forest, MLP 분류기에 각각 학습을 진행하였다. 데이터 규모가 제한적인 점을 고려하여 중첩 교차검증(Nested cross-validation)으로 일반화 성능을 평가하였다. Fig. 4와 같이 외부 폴드에서 전체 데이터를 여러

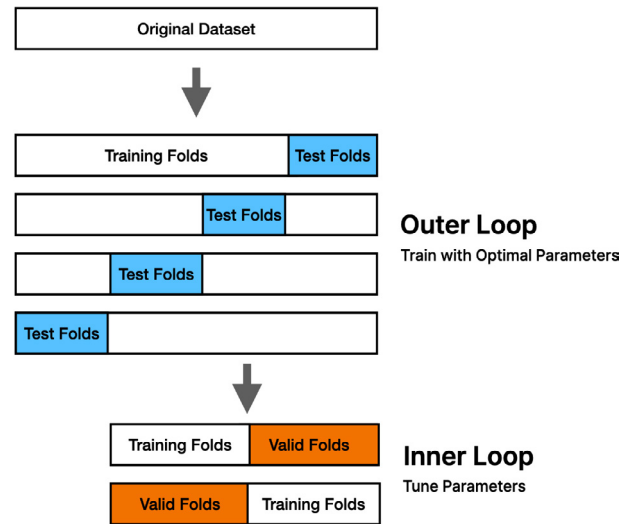


Fig. 4. Nested cross validation.

폴드로 분할하여 그중 하나를 테스트셋으로, 나머지를 학습셋으로 사용하여 모델의 최종 성능을 평가하였다. 내부 폴드에서는 외부 폴드의 학습셋만을 다시 여러 개의 폴드로 분할하여 GridSearchCV를 수행하여 수많은 조합 중 최적의 하이퍼파라미터를 탐색한다. 모델 간 비교는 아종 분류 성능 평가를 위해 정확도로 수행하였다. 학습을 진

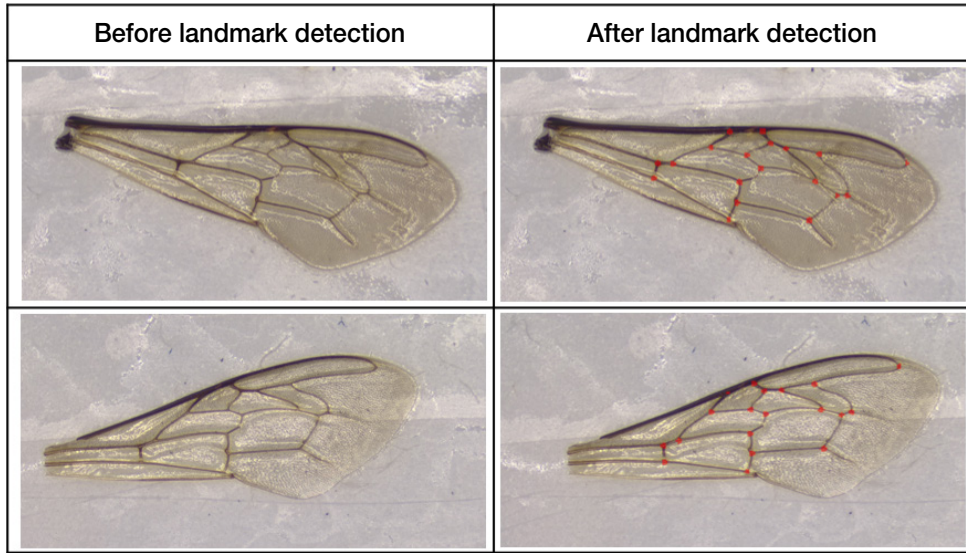


Fig. 5. Example of honey bee forewing images before and after landmark detection. The left panels show the original images, and the right panels show the detected landmark coordinates (red points) at wing vein intersections.

행한 환경은 하드웨어는 NVIDIA RTX 4090, 소프트웨어 환경은 Ubuntu 24.04, CUDA 12.4, cuDNN 8.9.4, Python 3.12.8, PyTorch 2.5.1, scikit-learn 1.6.1을 사용하였다.

결과 및 고찰

1. 시맥 교차점 탐지를 위해 U-Net 학습 후, 데이터세트 1을 이용한 추론 성능 평가

학습을 통해 얻은 최고의 가중치를 U-Net에 적용 후 데이터세트 1을 이용해 19개의 시맥 교차점 탐지 성능을 평가하였다(Fig. 5). Table 2는 학습세트, 검증세트의 IoU, Recall, Precision, Dice를 측정된 결과이다. 검증세트 기준으로 정밀도(Precision) 0.844, 재현율(Recall) 0.777로, 좌표 기반 판정에서 정밀도 우위가 확인된다. 중첩 기반 지표에서 Dice 0.809, IoU 0.682로 일관된 수준의 성능을 보였다. 학습세트와 검증세트 간 일반화 차이는 IoU -0.0444, Recall -0.0393, Precision -0.0240, Dice -0.0326으로 2.4~4.4%p 범위에 머물렀다.

2. 데이터세트 2를 이용하여 추론 후 변환된 특징을 이용한 아종 분류 평가

학습을 통해 얻은 best.pth를 U-Net에 적용하여 데이터

Table 2. Evaluation of U-Net model for landmark detection

| Dataset | Precision | Recall | Dice | IoU |
|---------|-----------|--------|-------|-------|
| Train | 0.868 | 0.816 | 0.868 | 0.726 |
| Valid | 0.844 | 0.777 | 0.844 | 0.682 |

Table 3. Accuracy evaluation of classification model for subspecies classification by nested-cross validation

| Classification model | LDA | SVM | Random Forest | MLP |
|----------------------|------|------|---------------|------|
| Average accuracy (%) | 86.4 | 86.3 | 77.8 | 85.2 |

세트 2를 이용해 각각의 분류기를 통해 학습 후 아종 분류 성능을 평가하였다. Table 3은 데이터세트 #2를 이용하여 학습된 분류기별 평가 결과이다. 교차 검증을 통한 성능 측정 결과는 LDA가 86.4%, SVM은 86.3%, Random Forest는 77.8%, MLP가 85.2%로 평가되었다. 실험 결과 LDA, SVM, MLP가 우수한 성능을 보여주었다.

3. 기계학습(자동 분류)을 통한 날개 시맥 교차점

자동 분류기를 통한 날개 시맥 교차점 분석의 경우, 국내에서 수집한 서양종 꿀벌 5개 계통의 좌표값을 기반으로 서열화 분석을 수행한 결과, Fig. 6과 같은 분포 양상을 보였다. 계통 간 분포는 비교적 명확하게 구분되었으며, A,

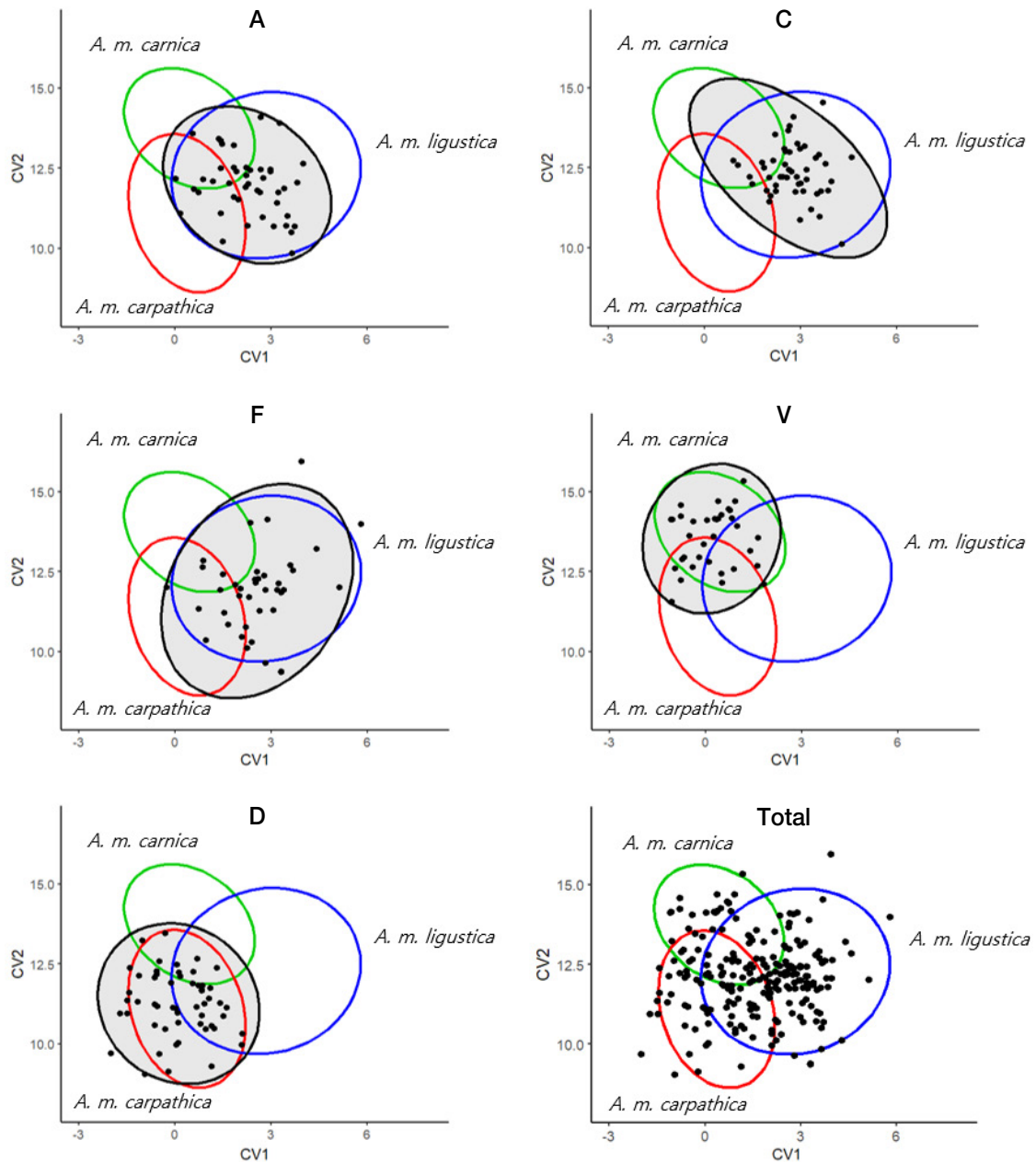


Fig. 6. Ordination plot of wing vein coordinates derived from machine learning-based landmark detection. The distribution of five honeybee lines (A, C, D, F, and V) is shown relative to reference subspecies regions (A = *ligustica*, C = *ligustica*, D = *carpathica*, F = *ligustica*, V = *carnica*).

C, F 계통은 주로 *ligustica* 영역을 중심으로 분포하는 경향을 보였다. 반면, D 계통은 *carpathica*, V 계통은 *carnica* 영역에 해당하는 위치에 분포하는 것으로 확인되었다.

4. 날개 시맥 형태 분류를 통한 아종 분석(수동 분류)

IdentiFly 프로그램으로 산출한 좌표값을 기반으로 서열화 분석을 진행하였다. 이때 사용된 개체는 자동 분류

기 분석에 사용된 시맥 데이터와 동일하였으며, 각 계통별 사용 개체 수는 Table 1에 제시하였다. 그 결과, 수동 시맥 교차점 기반 분석에서는 자동 분류기 결과에 비해 계통 간 분포의 분리가 상대적으로 명확하지 않았다. 특히 동일한 남·동유럽 계열에 속하는 *carnica*와 *ligustica* 간 경계가 일부 중첩되어 구분이 모호한 경향을 보였다(Fig. 7).

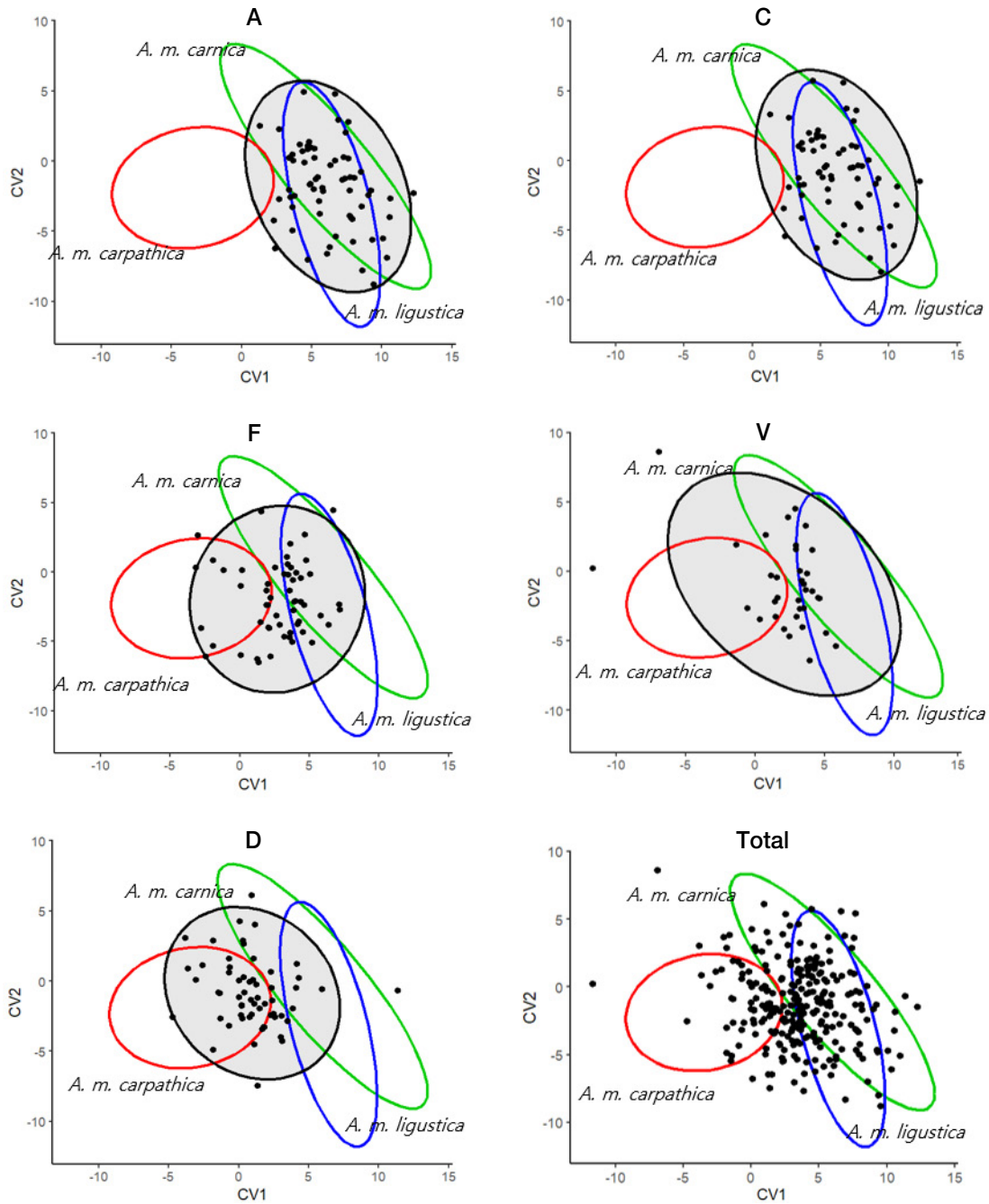


Fig. 7. Ordination plot of wing vein coordinates derived from manual detection with IdentiFly Program. The distribution of five honeybee lines (A, C, D, F, and V) is shown relative to reference subspecies regions (A = *ligustica*, C = *ligustica*, D = *carpathica*, F = *ligustica*, V = *carnica*).

5. 기계학습(자동 분류기)과 수동 분류기의 성능 비교

기계학습(자동 분류)을 이용한 날개 시맥 교차점 서열화 분석 결과, 모든 계통(A, C, D, F, V)에서 실제 아종과 일치하였다. 수동 분류의 경우 A, C, D, V 계통은 실제 아종

과 일치하였으나, F 계통에서는 실제로 *ligustica* 계통임에도 불구하고 *carnica*로 오분류되었다. 각 계통의 중심 좌표와 기준 아종의 중심 좌표 간 거리를 비교한 결과, 기계학습(자동 분류)은 모든 계통에서 실제 아종과의 거리가 가장 작게 나타났으며, 동시에 가장 가까운 다른 아종과의

Table 4. Comparison of classification accuracy and margin between automated and manual methods based on centroid distances in ordination space

| Line | Subspecies | Auto | | Manual | |
|--------------|-------------------|----------|--------|----------|--------|
| | | Accuracy | Margin | Accuracy | Margin |
| A | <i>ligustica</i> | ✓ | 1.867 | ✓ | 0.923 |
| C | <i>ligustica</i> | ✓ | 0.301 | ✓ | 0.172 |
| D | <i>carpathica</i> | ✓ | 2.208 | ✓ | 0.495 |
| F | <i>ligustica</i> | ✓ | 2.321 | ✗ | -0.246 |
| V | <i>carnica</i> | ✓ | 2.108 | ✓ | 0.906 |
| Accuracy (%) | | 100 | | 80 | |
| Mean margin | | 1.76 | | 0.45 | |

거리 차이(margin)가 상대적으로 크게 나타났다. 특히 D, F, V 계통에서는 정답 아종과 대체 아종 간 거리 차이가 크게 나타나, 분류 경계가 뚜렷하게 형성된 것을 확인할 수 있었다. 반면, 수동 분류에서는 전반적으로 정답 아종과 가장 가까운 대체 아종 간 거리 차이가 작게 나타났으며, 특히 *ligustica*를 기반으로 하는 A, C, F 계통에서는 정답 아종과 대체 아종 간 margin이 매우 작게 나타나 형태적 구분이 모호한 양상을 보였다(Table 4).

고 찰

본 연구는 국내에서 관리되고 있는 계통을 분석 대상으로 기계학습을 통해 형태형질에 기반한 아종 간 분류 가능성을 확인하였다. 시맥 교차점 지정 및 분류에서 수동 자료 추출보다 기계학습을 통한 자동 자료 추출의 정확도가 더 높게 나타났다. 또한 국내에서 관리되고 있는 서양꿀벌 계통(A, C, D, F, V)은 각 기원 아종이 특정되어 있으나, 본 연구 결과에서는 각 계통이 완전순계라고 보기는 어렵고 다양한 지역 기원의 아종이 혼재되어 나타났으며, 이는 도입 이전 집단에서 일정 수준의 유전적 혼합이 존재했을 가능성을 시사한다. 국내에서 관리되고 있는 A, C, D, F, V 계통은 2005~2008년 사이 국내에서 수집된 집단을 기반으로 순계 분리가 이루어진 것으로 보고되어 있으며(Lee et al., 2014, 2015), 분리 이전 집단에서 존재하던 형태적 또는 유전적 변이가 현재 계통에서 부분적으로 관찰될 가능성 또한 배제할 수 없다(Dilday, 2022).

수동 분류에 의한 아종 동정 결과를 바탕으로 각 계통의

기원별 분포 비율을 확인한 결과, 각 계통은 기원 아종 외에도 다른 계열 또는 아종으로 함께 분류되는 경향을 보였다. A 계통(*A. m. ligustica*)에서는 C 계열이 65%로 가장 높은 비율을 차지하였으나, O 계열 10%, A 계열 9%, 그리고 *A. m. carpathica* 16%로도 분류되었다. C 계통에서도 C 계열이 76%로 가장 높은 비율을 보였으나, A 계열 6%와 *A. m. carpathica* 16%가 함께 나타났다. D 계통에서는 *A. m. carpathica*가 50%로 가장 높은 비율을 차지하였으나, C 계열 26%와 A 계열 24%로도 분류되었다. F 계통에서는 C 계열이 63%로 가장 높은 비율을 보였으며, O 계열 4%, A 계열 9%, *A. m. carpathica* 24%가 함께 나타났다. V 계통 또한 C 계열이 66%로 가장 높은 비율을 차지하였으나, O 계열 10%, A 계열 18%, *A. m. carpathica* 6%로도 분류되었다.

또한 수동 랜드마크 기반 형태분석은 좌표 지정 과정에서 관찰자 의존적 오차가 발생할 수 있으므로, 본 연구에서는 이러한 개체 단위 오차의 영향을 최소화하기 위해 다수 개체를 대상으로 랜드마크를 지정하고 군집 단위의 평균 시맥 교차점 좌표를 이용하여 분석하였다. 형태적 동정 방법은 환경 조건 등에 의해 영향을 받을 수 있으므로(Oleksa and Tofilski, 2015), 날개 시맥 교차점 기반 분석뿐만 아니라 날개 길이, 체장, 흉폭, 두폭 등 다양한 형태형질을 포함한 통합적 형태측정 접근이 보다 신뢰성 높은 아종 분류를 위해 필요할 것으로 판단된다.

이와 함께 U-Net 기반 자동화 관점에서 볼 때, 본 연구의 의의는 단순히 분류 정확도를 제시하는 데에 그치지 않고, 기하학적 형태분석에서 가장 노동집약적인 단계인 시맥 교차점 지정 과정을 자동화하였다는 점에 있다. 기존의 수동 형태측정은 분석자의 숙련도와 판독 기준에 따라 결과가 달라질 수 있는 반면, 본 연구에서 제안한 U-Net 기반 파이프라인은 동일한 기준으로 대량의 시료를 반복적으로 처리할 수 있어 형태학적 분석의 재현성과 처리 효율을 동시에 향상시킬 수 있다. 실제로 수동 기반의 날개 시맥 형태측정은 개체 1마리를 분석하는 데 약 1분 정도가 소요되며, 100개체를 연속적으로 분석할 경우 약 1시간 40분이 필요하다. 또한 연구 목적의 분석에서는 더 많은 시료가 요구되므로 시간과 노동 부담이 더욱 증가하며, 분석자의 숙련도에 따라 소요 시간의 변동성도 크게 나타날 수 있다. 이와 비교할 때, 기계학습 기반 자동 분류는 대량의 시료를 신속하고 일관되게 처리할 수 있어 시간 및 노동 측

면에서 매우 높은 효율성을 제공하는 것으로 판단된다.

기계학습 기반 자동 분류는 효율성뿐만 아니라 정확도 측면에서도 수동 분류에 비해 뚜렷한 장점을 보였다. 본 연구에서 자동 분류 방법은 더 높은 정확도와 margin 값을 나타내어 꿀벌 아종 간 형태적 구분에 있어 우수한 성능을 보였다. 특히, 자동 분류는 모든 계통에서 일관되게 높은 정확도를 유지하였으며, 정답 아종과 인접 아종 간의 거리 차이가 크게 나타나 분류 경계가 보다 명확하게 형성되는 경향을 보였다. 반면, 수동 분류는 일부 계통에서 오분류가 발생하였으며, 정답 아종과 대체 아종 간 거리 차이가 작거나 음의 값을 보이는 경우도 확인되어 형태적 구분의 불확실성이 존재하였다. 이러한 결과는 수동 방식이 관찰자의 의존성과 랜드마크 설정의 주관성에 영향을 받을 수 있음을 시사한다. 본 연구에서는 관찰자 1인이 모든 랜드마크를 지정하였으나, 동일한 관찰자에 의해서도 좌표 지정 과정에서 관찰자 내 오차(observer-dependent measurement error)가 발생할 수 있어 형태적 구분의 불확실성이 나타날 수 있다(von Cramon-Taubadel *et al.*, 2007; Ridet *et al.*, 2020).

또한 본 연구에서 LDA와 SVM이 Random Forest보다 높은 성능을 보인 결과는, 제안한 210차원 특징 공간에서 집단 간 차이가 매우 복잡한 비선형 구조라기보다는 GPA 정렬과 거리 기반 특징을 통해 비교적 연속적이고 판별 가능한 구조로 표현되었을 가능성을 시사한다. 이는 날개 시맥 형태 정보가 적절한 정규화와 특징화 과정을 거칠 경우, 전통적 통계 기반 분류기와 마진 기반 분류기에서도 충분한 판별력을 가질 수 있음을 시사한다. 본 연구에는 몇 가지 한계가 존재한다. 첫째, 분류기 학습에 사용된 데이터셋이 비교적 소규모이며, 향후 대규모 데이터셋을 확보하여 분류 성능을 향상시킬 수 있을 것으로 기대된다. 둘째, 본 연구는 3개 아종만을 대상으로 하였으므로, 보다 다양한 아종을 포함한 추가 검증이 필요하다. 셋째, 촬영 조건이 단일 실제현미경으로 통일되어 있어, 다양한 촬영 환경에서의 모델 강건성(robustness)에 대한 검증이 요구된다. 향후 연구에서는 데이터 증강 및 촬영 조건 다변화를 통해 모델의 일반화 성능을 강화하고, 뒷날개 및 체형 특징을 통합한 다중 형질 기반 분류 파이프라인으로 확장할 필요가 있다. 종합하면, 본 연구의 U-Net 기반 자동 형태분석 파이프라인은 전통적인 기하학적 형태분석의 시간과 노동 부담을 줄이면서도 실용적인 분류 가능성을 보

여주었으며, 향후 양봉 현장에서 순계 관리와 기초 선별을 지원하는 보조적 도구로 활용될 가능성을 시사한다.

적 요

꿀벌의 아종 식별은 유전적 다양성 보존은 물론 양봉을 위한 육종 자원의 특징 이해를 위해 중요하다. 기존의 전통적 형태분석은 노동집약적이면서 정확도가 떨어진다. 분자유전학적 방법은 정확도는 높지만 비용이 많이 든다. 기하학적 형태분석법은 전통적 형태분석법에 비해 정확도가 높으며, 비용도 적게 든다. 그러나 19개의 시맥 교차점 수동 지정해야 하는 단점이 있다. 본 연구는 기하학적 형태분석법을 보완하여 앞날개 영상에서 19개 시맥 교차점 자동 탐지 - 형태 정규화 및 특징 변환 - 분류로 이어지는 파이프라인을 제안하고, 시맥 교차점 분할 모델과 지도 학습 분류기의 성능을 비교하였다. U-Net을 학습(데이터 세트 1)하여 최적 가중치(best.pth)로 추론하고, 헝가리안 매칭으로 템플릿 인덱스를 일치시켰다. GPA로 위치·스케일·회전 불변성을 확보한 뒤, GPA로 정렬된 좌표(38D)와 모든 점쌍 거리(171D), centroid size (1D)를 결합하여 210차원의 특징을 구성하였다. 중첩 교차검증으로 일반화 성능을 평가한 결과, 시맥 교차점 탐지는 검증세트 기준 Precision 0.844, Recall 0.777, Dice 0.844, IoU 0.682를 보였다. 약 400장으로 학습한 분류 성능은 LDA 86.4%, SVM 86.3%, MLP 85.2%, Random Forest 77.8%로 평가되었다. 제안된 파이프라인에서 U-Net 기반 자동 탐지는 실생활에서 사용가능한 정확도를 제공하며, 제안된 210차원 특징에서 선형 판별과 마진 기반 분류기가 트리 앙상블 기반 분류기보다 아종 분류 성능이 더 우수하다는 것을 확인하였다. 또한, 시맥 교차점 지정 과정을 자동화함으로써 시간 비용 측면에서도 기존 수동 분석에 비해 높은 효율성을 보였다.

사 사

이 논문은 국립경국대학교 기본연구지원사업에 의하여 연구되었음.

감사의 글

본 연구에 사용된 시료를 제공해 주신 김동원 박사님과 김진명 박사님께 깊이 감사드립니다.

인용 문헌

- Adam, J., E. D. Rothman, W. E. Kerr and Z. L. Paulino. 1977. Estimation of the number of sex alleles and queen matings from diploid male frequencies. *Genetics* 86: 583-596.
- Akongte, P. N., B. S. Park, M. Son, C. H. Lee, D. Oh, Y. S. Choi and D. Kim. 2024. The influence of environmental factors on site selection augment breeding success in honey bees: An insight of honey bee genetic resource conservation. *Biology* 13(6): 444.
- Chapman, N. C., B. A. Harpur, J. Lim, T. E. Rinderer, M. H. Allsopp, A. Zayed and B. P. Oldroyd. 2016. Hybrid origins of Australian honey bees (*Apis mellifera*). *Apidologie* 47: 26-34. <https://doi.org/10.1007/s13592-015-0371-0>
- Dilday, S. E. 2022. Using morphometric analysis of honey bee (*Apis mellifera*) colonies to identify lineages and subspecies on the isolated island of Newfoundland. Doctoral dissertation, Memorial University of Newfoundland, Canada.
- Dogantzis, K. A., T. Tiwari, I. M. Conflitti, A. Dey, H. M. Patch, E. M. Muli, L. Garnery, C. W. Whitfield, E. Stolle, A. S. Alqarni, M. H. Allsopp and A. Zayed. 2021. Thrive out of Asia and the adaptive radiation of the western honey bee. *Sci. Adv.* 7(49): eabj2151. <https://doi.org/10.1126/sciadv.abj2151>
- Francoy, T. M., D. Wittmann, M. Drauschke, S. Müller, V. Steinhage, M. A. F. Bezerra-Laure, D. De Jong and L. S. Goncalves. 2008. Identification of Africanized honey bees through wing morphometrics: two fast and efficient procedures. *Apidologie* 39(5): 488-494.
- Frunze, O., Y.-S. Choi, D.-W. Kim, B.-S. Park, H.-G. Park and E.-J. Kang. 2020. Morphometric characterization of honey bee, *Apis mellifera* Linnaeus, inbred lines in Korea. *J. Appl. Entomol.* 59(4): 371-382.
- Geldmann, J. and J. P. González-Varo. 2018. Conserving honey bees does not help wildlife. *Science* 359(6374): 392-393.
- Ilyasov, R. A., M.-L. Lee, J. I. Takahashi, H.-W. Kwon and A. G. Nikolenko. 2020. A revision of subspecies structure of western honey bee *Apis mellifera*. *Saudi J. Biol. Sci.* 27(12): 3615-3621.
- Jung, C. 2008. Economic value of honey bee pollination on major fruit and vegetable crops in Korea. *J. Apic.* 23: 147-152.
- Jung, C. 2022. Multi-functionality of honey bees for eco-friendly food production. *Food Sci. Ind.* 55(2): 166-175.
- Kandemir, I., A. Ozkan and S. Fuchs. 2011. Reevaluation of honeybee (*Apis mellifera*) microtaxonomy: A geometric morphometric approach. *Apidologie* 42: 18-627.
- Kang, Y., S. Kwon, B. A. Kakhramanov and C. Jung. 2025. Comparative analysis of morphological characteristics among subspecies of honey bee (*Apis mellifera*) in different regions of Eastern Uzbekistan. *J. Apic.* 40(2): 97-104.
- Kim, H.-K., C. G. Park and G.-I. Han. 2021. Evaluation of hygienic ability for the selection of disease-resistant honey bee (*Apis mellifera*) lines. *J. Bio-Env. Con.* 30(3): 206-211.
- Kim, H.-K., M.-L. Lee, M.-Y. Lee, Y.-S. Choi, D. Kim and A. R. Kang. 2017. Identification of a single nucleotide polymorphism (SNP) marker for the detection of enhanced honey production in honey bee. *J. Apic.* 32(3): 147-154.
- Kim, H.-K., M.-Y. Lee and M.-L. Lee. 2015. Hygienic behavior test of six inbred lines in *Apis mellifera* using the freeze-killed brood method. *J. Apic.* 30: 187-190.
- Klein, A. M., B. E. Vaissière, J. H. Cane, I. Steffan-Dewenter, S. A. Cunningham, C. Kremen and T. Tscharntke. 2007. Importance of pollinators in changing landscapes for world crops. *Proc. R. Soc. B Biol. Sci.* 274(1608): 303-313.
- Lee, M.-L., M.-Y. Lee, H.-S. Sim, Y.-S. Choi, H.-K. Kim and G.-H. Byoun. 2014. Characteristics of superior triple-crossed honey bee (*Apis mellifera* L.): Honey collection, hibernation, and hygienic behavior. *J. Apic. Res.* 29: 257-262.
- Lee, M.-L., M.-Y. Lee, Y.-S. Choi and H.-K. Kim. 2015. Discrimination of European honey bee (*Apis mellifera* L.) lines in Korea based on wing vein characters and molecular markers. *J. Apic.* 30(4): 245-251.
- Lee, M. R., Y.-S. Choi, D.-W. Kim and M.-Y. Lee. 2019. Age-dependent hypopharyngeal gland development and morphometric characteristics in crossbred honey bee lineages reared for high royal jelly production. *J. Asia-Pac. Entomol.* 22: 699-704.
- Lee, M.-Y., H.-K. Kim, M.-L. Lee, Y.-S. Choi, S.-M. Han and D.-W. Kim. 2017. Comparison of royal jelly production among crossbred honey bees during nectar flow and non-nectar flow periods. *J. Apic.* 32: 385-389.
- Mackensen, O. 1955. Further studies on a lethal series in the honey bee. *J. Hered.* 46: 72-74.
- Meixner, M. D., M. A. Pinto, M. Bouga, P. Kryger, E. Ivanova and S. Fuchs. 2013. Standard methods for characterising subspecies and ecotypes of *Apis mellifera*. *J. Apic. Res.* 52(4): 1-28.
- Miguel, I., M. Baylac, M. Iriondo, C. Manzano, L. Garnery and A. Estonba. 2011. Both geometric morphometric and microsatellite data consistently support the differentiation of the *Apis mellifera* M evolutionary branch. *Apidologie* 42(2): 150-161.
- Moritz, R. F. A., S. Hartel and P. Neumann. 2005. Global invasions of the western honey bee (*Apis mellifera*) and the

- consequences for biodiversity. *Ecoscience* 12(3): 289-301.
- Muñoz, I. and P. De la Rúa. 2021. Wide genetic diversity in Old World honey bees threaten by introgression. *Apidologie* 52(1): 200-217.
- Nawrocka, A., Í. Kandemir, S. Fuchs and A. Tofilski. 2018. Computer software for identification of honey bee subspecies and evolutionary lineages. *Apidologie* 49(2): 172-184.
- Oleksa, A. and A. Tofilski. 2015. Wing geometric morphometrics and microsatellite analysis provide similar discrimination of honey bee subspecies. *Apidologie* 46(1): 49-60.
- Pasupuleti, V. R., L. Sammugam, N. Ramesh and S. H. Gan. 2017. Honey, propolis, and royal jelly: A comprehensive review of their biological actions and health benefits. *Oxid. Med. Cell. Longev.* 2017: 1259510.
- Potts, S. G., V. Imperatriz-Fonseca, H. T. Ngo, M. A. Aizen, J. C. Biesmeijer, T. D. Breeze and A. J. Vanbergen. 2016. Safeguarding pollinators and their values to human well-being. *Nature* 540(7632): 220-229.
- Ridel, A. F., F. Demeter, M. Galland, E. N. L'abbe, D. Vandermeulen and A. C. Oettle. 2020. Automatic landmarking as a convenient prerequisite for geometric morphometrics. Validation on cone beam computed tomography (CBCT)-based shape analysis of the nasal complex. *Forensic Sci. Int.* 306: 110095.
- Roffet-Salque, M., M. Regert, R. P. Evershed, A. K. Outram, L. J. E. Cramp, O. Decavallas, J. Dunne, P. Gerbault, S. Mileto, S. Mirabaud, M. Pääkkönen, J. Smyth, L. Šoberl, H. L. Whelton, A. Alday-Ruiz, H. Asplund, M. Bartkowiak, E. Bayer-Niemeyer, L. Belhouchet, F. Bernardini, M. Budja, G. Cooney, M. Cubas, E. M. Danaher, M. Diniz, L. Domboróczki, C. Fabbri, J. E. González-Urquijo, J. Guilaine, S. Hachi, B. N. Hartwell, D. Hofmann, I. Hohle, J. J. Ibáñez, N. Karul, F. Kherbouche, J. Kiely, K. Kotsakis, F. Lueth, J. P. Mallory, C. Manen, A. Marciniak, B. Maurice-Chabard, M. A. McGonigle, S. Mulazani, M. Özdoğan, O. S. Perić, S. R. Perić, J. Pétrasch, M. Pétrequin, P. Petrequin, U. Poensgen, C. J. Pollard, F. Poplin, G. Radi, P. Stadler, H. Stäuble, N. Tesić, D. Urem-Kotsou, J. B. Vuković, F. Walsh, A. Whittle, S. Wolfram, L. Zapata-Peña and J. Zoughlami. 2015. Widespread exploitation of the honeybee by early Neolithic farmers. *Nature* 527: 226-230.
- Ronneberger, O., P. Fischer and T. Brox. 2015. U-Net: Convolutional networks for biomedical image segmentation. in *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015)*, eds. by Navab, N., J. Hornegger, W. Wells, A. Frangi. *Lecture Notes in Computer Science*, Vol. 9351, pp. 234-241. Springer, Cham.
- Ruttner, F. 1988. *Biogeography and taxonomy of honey bees*. Springer-Verlag, Berlin.
- Ruttner, F. 2013. *Biogeography and taxonomy of honey bees*. Springer Science & Business Media, Dordrecht.
- Tofilski, A. 2008. Using geometric morphometrics and standard morphometry to discriminate three honeybee subspecies. *Apidologie* 39: 558-563.
- von Cramon-Taubadel, N., B. C. Frazier and M. M. Lahr. 2007. The problem of assessing landmark error in geometric morphometrics: Theory, methods, and modifications. *Am. J. Phys. Anthropol.* 134(1): 24-35.
- Whitfield, C. W., S. K. Behura, S. H. Berlocher, A. G. Clark, J. S. Johnston, W. S. Sheppard, D. R. Smith, A. V. Suarez, D. Weaver and N. D. Tsutsui. 2006. Thrice out of Africa: Ancient and recent expansions of the honey bee, *Apis mellifera*. *Science* 314: 642-645.
- Woyke, J. 1976. Population genetics studies on sex alleles in the honey bee using the example of the Kangaroo Island bee sanctuary. *J. Apic. Res.* 15: 105-123.