### **1. Analiza wyników**

#### **Dotrenowany model: ViT (google/vit-base-patch16-224)**

* **Precyzja makro:** 0.9610
* **Recall makro:** 0.9597
* **F1-Score makro:** 0.9572
* **Dokładność:** 96.07%
* **Czas trwania testowania:** 275.5259 sekundy

Model ViT, będący dotrenowanym modelem, osiągnął doskonałe wyniki we wszystkich metrykach. Wysoka precyzja, recall oraz F1-score wskazują, że model skutecznie radzi sobie z klasyfikacją nawet w przypadku nierównomiernie rozłożonych danych między klasami.

#### **model: ResNet-50**

* **Precyzja makro:** 0.01
* **Recall makro:** 0.02
* **F1-Score makro:** 0.01
* **Dokładność:** 2.34%

Model ResNet-50 bez dodatkowego treningu na danych roślinnych całkowicie zawiódł. Uzyskane wyniki są zbliżone do losowego zgadywania, co wskazuje na brak dostosowania modelu do specyfiki danych (w naszym wypadku obrazów roślin o szczególnych cechach wizualnych).

#### **SwinV2-tiny-patch4window16-256**

* **Precyzja makro:** 0.7241
* **Recall makro:** 0.2069
* **F1-Score makro:** 0.0000
* **Czas trwania testowania:** 171.3213 sekundy

Model SwinV2 osiągnął umiarkowaną precyzję, co wskazuje, że czasami poprawnie identyfikował klasy, jednak bardzo niska wartość recall i F1-score pokazuje, że miał trudności z rozpoznawaniem większości klas.

#### **apple/mobilevit-small**

* **Precyzja makro:** 0.6250
* **Recall makro:** 0.3750
* **F1-Score makro:** 0.0000
* **Dokładność:** 0.00%
* **Czas trwania testowania:** 103.7489 sekundy

MobileViT uzyskał wyniki, które sugerują słabą wydajność w tym zadaniu. Brak dotrenowania na danych specyficznych dla roślin spowodował, że model praktycznie nie potrafił rozpoznać żadnej klasy.

### **2. Wyjaśnienie różnic między modelami**

1. **Dotrenowany model (ViT)**
   * Model Vision Transformer (ViT) został dotrenowany na danych specyficznych dla roślin, co pozwoliło mu dostosować swoje reprezentacje wewnętrzne do cech kluczowych dla tego zadania (np. kształtu liści, kolorów, tekstur).
2. **Modele bez dotrenowania (ResNet-50, SwinV2, MobileViT)**
   * Modele te zostały jedynie wstępnie wytrenowane na zbiorach danych ogólnych (np. ImageNet), które zawierają obrazy z różnych kategorii, ale niezwiązanych bezpośrednio z roślinami.
   * Brak adaptacji do specyfiki danych testowych skutkował bardzo słabymi wynikami, ponieważ cechy wizualne roślin mogą różnić się znacząco od obrazów z ogólnych zbiorów danych.
3. **Różnica w architekturach**
   * ViT jako model transformatorowy jest bardziej skuteczny w wykrywaniu subtelnych cech na obrazach, co czyni go odpowiednim do klasyfikacji złożonych danych wizualnych, takich jak rośliny.
   * ResNet i SwinV2, mimo że dobrze radzą sobie z ogólnymi zadaniami, mogą mieć trudności w identyfikacji bardzo podobnych klas, jeśli nie zostały dostosowane do specyficznego zadania.
   * MobileViT, jako lekka architektura zoptymalizowana pod kątem urządzeń mobilnych, ma ograniczoną zdolność do rozpoznawania szczegółowych cech w danych bez dodatkowego treningu.

Model testowany na zbiorze test wybraliśmy metryki makro ze względu na sporą nierównomierny rozkład danych w każdej z klas.