

Kategoryzacja chorób jabłoni na podstawie wyglądu ich liści

Mateusz Garczyński, Zuzanna Twardowska,
Wiktor Wołek, Dawid Wysocki



Rysunek 1: Choroby widoczne na liściach jabłoni. Źródło: kaggle [1].

1 Krótki opis projektu

Celem projektu jest wykrywanie chorób jabłoni na podstawie kolorowych zdjęć liści jabłoni.

Zbadanie stanu zdrowia liścia zostanie przeprowadzone w dwóch etapach:

1. Wyodrębnienie liścia z obrazu
2. Analiza liścia pod kątem obecności choroby

2 Wstępna analiza danych

Podczas pierwszej fazy pracy nad projektem, jaką była wstępna analiza dostępnych danych, skupiliśmy się na wyodrębnieniu jak największej liczby elementów wspólnych dla kolejnych zdjęć przedstawiających liście jabłoni.

2.1 Wyodrębnienie liścia z obrazu

Jako, że naszym pierwszym celem jest wykrycie badanego liścia na zdjęciu, to właśnie temu elementowi analizy poświęciliśmy najwięcej uwagi. Naszym pierwotnym pomysłem była analiza kolorystyczna zdjęcia, jednak ze względu na

fakt, że w tle na zdjęciach obecne są głównie inne liście niż badany obiekt (duże szумы na zdjęciach), doszliśmy do wniosku, że wyodrębnienie konturu liścia dzięki analizie kolorystycznej nie jest najbardziej optymalnym rozwiązaniem.

Drugim nasuwającym nam się pomysłem jest wyodrębnienie liścia za pomocą jego ostrości na zdjęciu - na większości zdjęć badany liść jest jedynym ostrym elementem zdjęcia. Tym samym, przy pomocy odpowiedniego "preprocessingu", spróbujemy wykorzystać zaobserwowaną zależność ostrości do pozycji badanego liścia w celu wyodrębnienia go z całości zdjęcia.

Dodatkowo chcemy wykorzystać podejście "trust the net" i zobaczyć na ile skuteczna będzie sieć wykrywająca badany element bez skupiania się na ostrych fragmentach zdjęcia, a jedynie ucząca się wykrywać odpowiednie kontury. Będzie to pierwsze z zastosowanych przez nas możliwych rozwiązań.

2.2 Analiza liścia pod kątem obecności choroby

Po wyodrębnieniu liścia ze zdjęcia należy zbadać, czy jest on zdrowy czy chory, a w przypadku wykrycia choroby przyporządkować liść do danej choroby lub grupy chorób. Podczas przeglądania zebranych zdjęć zauważyliśmy, że choroby liści opierają się przede wszystkim na ich punktowych zmianach kolorystycznych. W związku z zaobserwowaniem tego faktu, stan zdrowia liści będziemy sprawdzać przez analizę kolorów liścia i sprawdzeniu, czy istnieją skupiska o kolorze różniącym się od jego typowego koloru, a także przez zliczanie ilości i wielkości punktów podejrzanych o bycie symptomem choroby.

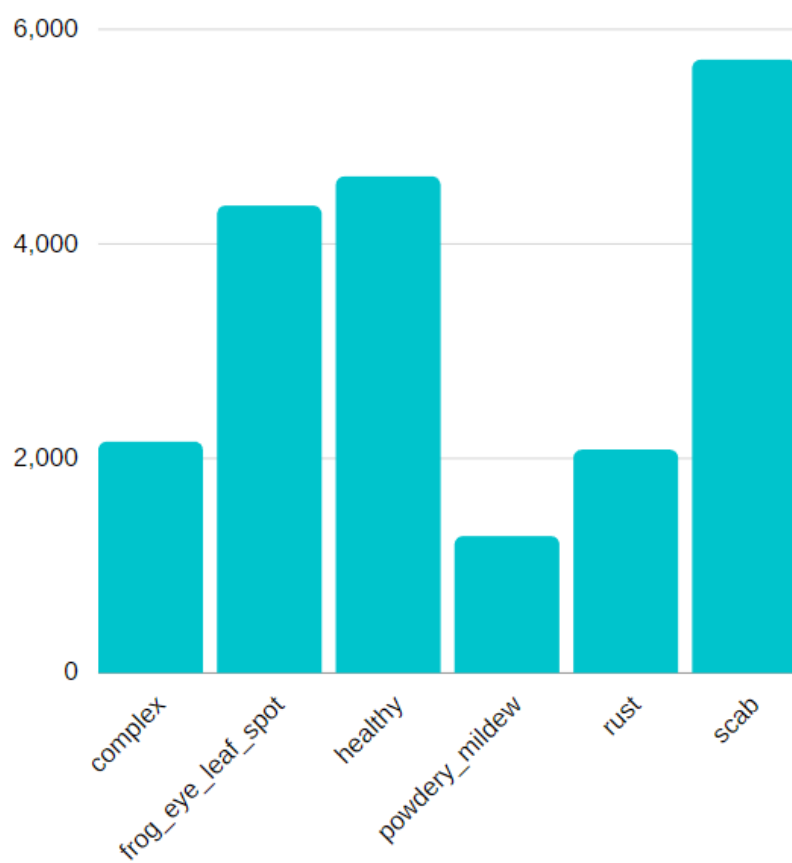
Inne podejście jednak będzie konieczne podczas analizy liścia pod kątem jednej z chorób, która objawia się całościowym pokryciem liścia lekko szarawym nalotem.

3 Rozkład danych

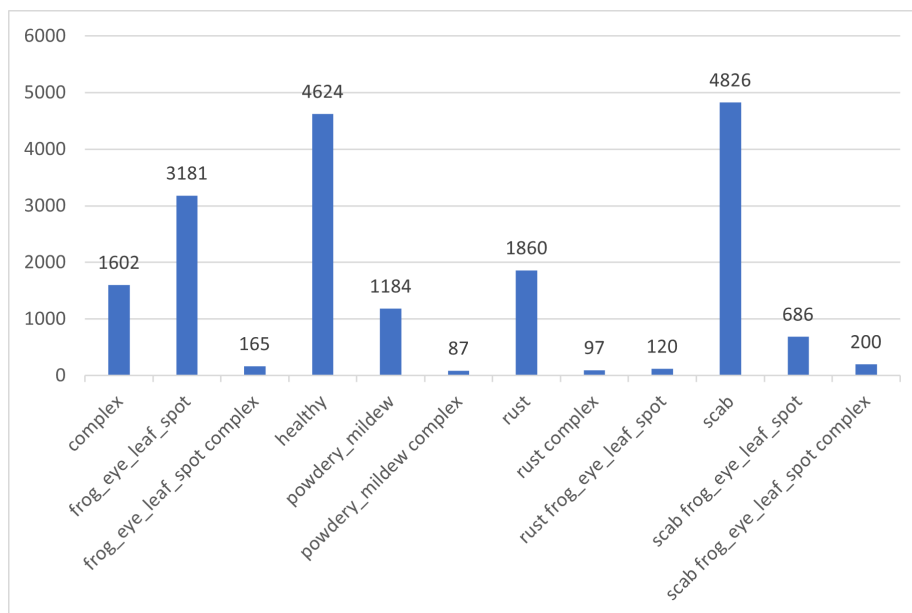
Na rysunkach 2 i 3 zaprezentowany został podział danych na kategorie odpowiadające chorobom. W zbiorze rozpoznajemy 5 klas chorób: "powdery mildew", "rust", "scab", "frog eye leaf spot" oraz "complex" oznaczającą zdjęcia, na których jest zbyt wiele chorób by zidentyfikować każdą pojedynczą. Dodatkowo wyróżniamy klasę zdrowych "healthy". Warto zauważyć, że jednocześnie jeden liść może posiadać kilka chorób. Przykładowe zdjęcia ze zbioru zostały przedstawione na rysunkach 4, 5, 6, 7, 8 i 9.

Rysunek 2 przedstawia podliczenie zdjęć liści od ich choroby. Jeśli na zdjęciu zostały wykryte dwie choroby, będziemy zaliczać je do dwóch klas chorób. Analogicznie postąpimy przy większej ilości chorób. Kategoria "complex" reprezentuje zdjęcia, na których jest zbyt wiele chorób aby zidentyfikować pojedyncze z nich.

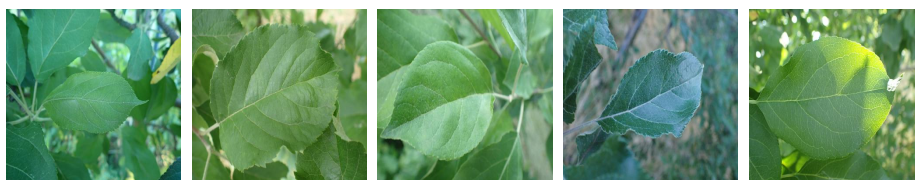
Rysunek 3 przedstawia podział na 12 klas zawierający wszystkie kombinacje chorób zawartych w zbiorze danych z kaggle. Możemy zauważyć, że niektóre klasy np "rust complex" mają niewielką licznosc, co utrudnia trenowanie modelu.



Rysunek 2: Liczba zdjęć z daną chorobą.



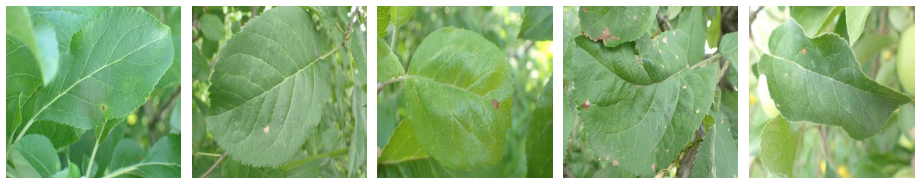
Rysunek 3: Liczności zdjęć liści z kombinacjami chorób do nich przypisanymi.



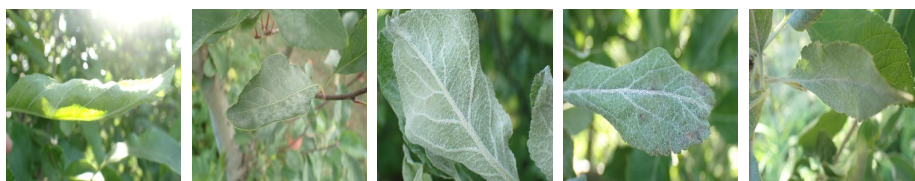
Rysunek 4: Przykładowe zdjęcia ze zbioru przedstawiające zdrowe jabłonie (healthy).



Rysunek 5: Przykładowe zdjęcia ze zbioru przedstawiające jabłonie ze znaczącą ilością chorób (complex).



Rysunek 6: Przykładowe zdjęcia ze zbioru przedstawiające jabłonie z botryosphaeria obtusa (frog_eye_leaf_spot).



Rysunek 7: Przykładowe zdjęcia ze zbioru przedstawiające jabłonie z mączniakiem prawdziwym (powdery_mildew).



Rysunek 8: Przykładowe zdjęcia ze zbioru przedstawiające jabłonie z parchem (scab).



Rysunek 9: Przykładowe zdjęcia ze zbioru przedstawiające jabłonie posiadające rdzowce (rust).

4 Szczegółowy plan zadań

Na obecnym etapie wydzieliliśmy następujący podział zadań:

- Mateusz Garczyński - zapoznanie się z dostępnymi narzędziami do object recognition, analiza wyników gotowego programu
- Zuzanna Twardowska - zapoznanie się z artykułami naukowymi dotyczącymi analizy kolorystycznej zdjęć, analiza artykułów dotyczących pre-processingu
- Wiktor Wołek - podział zdjęć na foldery (konieczne dla pracy z Keras), przygotowanie histogramu obrazującego podział zdjęć na kategorie odpowiadające chorobom, zmniejszenie zdjęć do odpowiedniego rozmiaru, analiza wyników gotowego programu
- Dawid Wysocki - zapoznanie się z artykułami naukowymi dotyczącymi analizy zdjęć pod kątem ostrych fragmentów, analiza dostępnych platform obliczeniowych i możliwości swojego komputera

Literatura

- [1] Platforma kaggle.com: <https://www.kaggle.com/c/plant-pathology-2021-fgvc8/>