

Zastosowanie zagnieżdżonej dychotomii w klasyfikacji wieloklasowej owoców i warzyw

Tomasz Michalski i Joanna Orzeł

Politechnika Wrocławska, wybrzeże Stanisława Wyspiańskiego 27, 50-370 Wrocław,
Polska

Abstract. Problemy multiklasowe są popularnym scenariuszem w problemie klasyfikacji owoców i warzyw. Skutecznym podejściem do rozwiązania tego problemu jest zastosowanie zagnieżdżonej dychotomii. W poniższym badaniu na przygotowanym modelu zagnieżdżonej dychotomii zostały przeprowadzone eksperymenty, które miały na celu znalezienie najlepszej metody selekcji cech, balansowania danych i hiperparametrów drzew decyzyjnych. Zostały także porównane dwie metody zagnieżdżonej dychotomii, pierwsza metoda w której dane z klas poprzednio uczonych są usuwane i druga metoda w której każdy węzeł jest uczony na wszystkich danych.

Keywords: Zagnieżdżonadychotomia · Klasyfikacja wieloklasowa · Drzewa decyzyjne.

1 Temat Projektu

1.1 Wprowadzenie

W obecnych czasach klasyfikacja obiektów poprzez ich zdjęcia to nieodłączny element dziedzin takich jak uczenie maszynowe i wizja komputerowa. Zaawansowane algorytmy i techniki pozwalają na coraz bardziej precyzyjne rozpoznawanie i kategoryzowanie obiektów, co znajduje szerokie zastosowanie w różnych branżach i dziedzinach życia.

Klasyfikator wieloklasowy w przeciwieństwie do klasyfikatora binarnego powinien klasyfikować obiekty na więcej niż 2 klasy. Jest to bardziej skomplikowany proces przysparzających wielu problemów takich jak brak balansu danych, problemy z przygotowaniem danych do treningu czy niewystarczająca ilość danych.

Jedną z metod mającą usprawnienie procesu klasyfikacji wieloklasowej jest zagnieżdżona dychotomia. Metoda ta polega na stworzeniu zespołu klasyfikatorów, gdzie każdy jest klasyfikatorem binarnym. Jako zespół są w stanie klasyfikować wiele klas. Atutem tego podejścia jest większa ilość danych do treningu dla każdego klasyfikatora binarnego, czyli węzła oraz optymalizacja czasu potrzebnego do wytrenowania modelu.

W dalszej sekcji dokumentu zaprezentowany jest przegląd poprzednich prac w temacie zagnieżdżonej dychotomii. Sekcja druga zawiera opis proponowanej

metody zagnieżdżonej dychotomii oraz kolejności tworzenia węzłów poszczególnych klas. W dalszej części opisane zostały przeprowadzone eksperymenty w których sprawdzane są metody wyodrębniania cech, rodzaje implementacji zagnieżdżonej dychotomii, metody balansowania danych oraz doборы parametrów. Końcowo w sekcji trzeciej uwzględniono wyniki oraz wnioski z przeprowadzanych badań.

1.2 Studia literaturowe

Zagnieżdżona dychotomia jest zaawansowaną techniką klasyfikacji, która pozwala na efektywniejsze radzenie sobie z problemami wieloklasowymi poprzez podział na mniejsze, bardziej zarządzalne podproblemy binarne. W kontekście klasyfikacji obrazów, technika ta może być szczególnie przydatna w automatyzacji procesów sortowania i oceny jakości produktów rolnych, co ma bezpośredni wpływ na bezpieczeństwo żywnościowe i efektywność produkcji[1].

Wyodrębnianie cech obrazowych odgrywa kluczową rolę w procesie klasyfikacji. Modele takie jak VGG16, InceptionV3 i MobileNetV2 są szeroko stosowane w dziedzinie uczenia maszynowego ze względu na ich zdolność do ekstrakcji bogatych reprezentacji cech z obrazów. Te głębokie sieci neuronowe, dzięki swoim architekturom, umożliwiają identyfikację i klasyfikację różnorodnych gatunków owoców i warzyw z wysoką dokładnością[14,15,16,17].

Metody balansowania danych są niezbędne w przypadkach, gdy mamy do czynienia z niezbalansowanymi zbiorami danych, co jest typowe dla problemów klasyfikacji wieloklasowej. Techniki takie jak RandomOverSampler, SMOTE, RandomUnderSampler, TomekLinks oraz SMOTETomek pozwalają na lepsze zarządzanie proporcjami klas, co przekłada się na poprawę wydajności modeli klasyfikacyjnych[2,3,4,5].

Optymalizacja parametrów modeli jest kolejnym kluczowym elementem, który może znacząco wpłynąć na skuteczność klasyfikacji. Metody takie jak GridSearchCV, RandomizedSearchCV i BayesSearchCV pozwalają na systematyczne przeszukiwanie przestrzeni parametrów w celu znalezienia optymalnej konfiguracji dla danego modelu[6,7,8,9,10,11,12,13].

Warto zauważyć, że zastosowanie zagnieżdżonej dychotomii w połączeniu z zaawansowanymi technikami przetwarzania obrazów i metodami balansowania danych, wspierane przez skuteczne strategie optymalizacji parametrów, stanowi potężne narzędzie w dziedzinie klasyfikacji wieloklasowej owoców i warzyw. Kontynuacja badań w tym obszarze może przynieść nowe innowacje i ulepszenia, które będą miały znaczący wpływ na przyszłość przetwarzania obrazów i uczenia maszynowego.

W kontekście badań nad zagnieżdżoną dychotomią, istotne jest również odniesienie do istniejących prac naukowych. Przykładowo, praca Michalskiej-Ciekańskiej (2022) dotycząca klasyfikacji zmian skórnych wskazuje na znaczenie wyboru odpowiedniej bazy danych, wstępnego przetwarzania obrazów, doboru zestawu danych uczących, narzędzi klasyfikacji, etapu trenowania sieci i otrzymania wyników końcowych w procesie klasyfikacji wieloklasowej[18]. Podobnie, praca Koszeli i Weresa (2005) prezentuje metodę analizy i klasyfikacji

obrazów suszu warzywnego z wykorzystaniem sztucznych sieci neuronowych, co pokazuje potencjał zastosowania zaawansowanych technik przetwarzania obrazów w rolnictwie[1].

Podsumowując, zagnieżdżona dychotomia w połączeniu z zaawansowanymi technikami przetwarzania obrazów i metodami balansowania danych, wspierana przez skuteczne strategie optymalizacji parametrów, stanowi potężne narzędzie w dziedzinie klasyfikacji wieloklasowej owoców i warzyw. Kontynuacja badań w tym obszarze może przynieść nowe innowacje i ulepszenia, które będą miały znaczący wpływ na przyszłość przetwarzania obrazów i uczenia maszynowego.

2 Przygotowanie do eksperymentu

2.1 Opis problemu i algorytmiki

Celem tego badania jest opracowanie modelu uczenia maszynowego, który będzie w stanie klasyfikować owoce i warzywa na podstawie ich zdjęć (cech wizualnych). Dane zawierają zdjęcia owoców i warzyw oraz ich etykiety z nazwą. Zmienne w modelu to:

- Zdjęcia pojedynczego owocu lub warzywa, reprezentowane jako macierz pikseli
- Etykiety klasy - zmienna określająca rodzaj produktu na zdjęciu, np. "jabłko", "banan" itd.

Funkcją celu jest entropia krzyżowa, obliczana pomiędzy prawdziwymi etykietami klasowymi a przewidywanymi przez model. Celem jest minimalizacja tej funkcji podczas trenowania modelu. Entropia krzyżowa mierzy odległość pomiędzy dwoma rozkładami prawdopodobieństwa: rzeczywistymi etykietami klas, a przewidywanymi przez model. Im bliższe są wartości rozkładów prawdopodobieństwa do siebie, tym niższa wartość entropii krzyżowej, co oznacza lepsze dopasowanie modelu do rzeczywistości. Prawdopodobieństwo, które jest używane w entropii krzyżowej, pochodzi z przewidywań modelu. Model przypisuje każdemu obiektowi pewne prawdopodobieństwo przynależności do poszczególnych klas. Te przewidywane prawdopodobieństwa mogą być interpretowane jako stopień pewności modelu do której klasy należy dany obiekt.

Obrazy przy użyciu gotowych sieci neuronowych zostają zamienione na wektory cech, które są przystępniejsze do uczenia modelu. Podejście, w którym używa się wcześniej wytrenowanych modeli nazywa się transfer learning. Dane poprzez 5-krotną walidację krzyżową są dzielone na zbiory treningowe i walidacyjne. Model używa struktury zagnieżdżonej dychotomii do podziału klas na mniejsze problemy (zbiory klas). W węzłach użyte są drzewa decyzyjne do klasyfikacji binarnej.

Zagnieżdżona dychotomia jest techniką hierarchicznego podziału problemu klasyfikacji wieloklasowej na mniejsze podproblemy binarne. Algorytm zagnieżdżonej dychotomii działa w sposób, gdzie każdy węzeł uczony jest rozróżniać jedną klasę od innych.

Kolejność tworzenia węzłów dla poszczególnych klas obliczana jest poprzez indeks Giniego, który używa drzew decyzyjnych do oceny ważności cech. Indeks Giniego w kontekście uczenia maszynowego, szczególnie w drzewach decyzyjnych, jest miarą nieczystości węzła. Mierzy on prawdopodobieństwo, że przypadkowo wybrany element zostanie nieprawidłowo sklasyfikowany, gdyby jego etykieta została losowo przydzielona zgodnie z rozkładem klas w danym węźle.

$$Gini(t) = 1 - \sum_{i=1}^n (p_i)^2 \quad (1)$$

gdzie:

- $Gini(t)$ to wartość indeksu Giniego dla węzła t ,
- n to liczba klas
- p_i o proporcja przykładów należących do klasy i

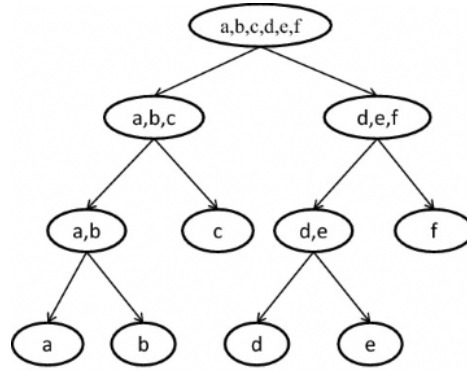


Fig. 1. Przykład rozkładu klas w zagnieżdżonej dychotomii

Funkcja zagnieżdżonej dychotomii została zdefiniowana według wzoru 2.

$$f(x) = \begin{cases} y_i & \text{jeśli liczba klas } y_i \leq 1 \\ f_{\text{lewy}}(x) & \text{jeśli cecha } x \text{ spełnia warunek podziału} \\ f_{\text{prawy}}(x) & \text{jeśli cecha } x \text{ nie spełnia warunku podziału} \end{cases} \quad (2)$$

gdzie:

- y_i to etykieta klasy dla obiektu x ,
- $f_{\text{lewy}}(x)$ i $f_{\text{prawy}}(x)$ to wywołania funkcji dla lewej i prawej podgrupy obiektów.

Proces podziału kontynuuje się, aż do uzyskania liści reprezentujących wszystkie klasy.

Zagnieżdżona dychotomia będzie używana do hierarchicznego podziału różnych gatunków owoców i warzyw na podklasy binarne, co umożliwi bardziej skuteczną klasyfikację.

Drzewa decyzyjne to algorytmy, które pozwalają na podejmowanie decyzji na podstawie pytań o cechy obiektów. Podczas uczenia drzewa decyzyjnego, algorytm szuka najlepszej cechy podziału w każdym węźle, która najlepiej separuje obiekty należące do różnych klas. Gałęzie w drzewie decyzyjnym reprezentują możliwe odpowiedzi do pytania w węźle.

Proces tworzenia drzewa decyzyjnego może być opisany za pomocą funkcji podziału $\text{split}(X, y)$, która wybiera najlepszą cechę podziału na podstawie zbioru obiektów X i odpowiadających im etykiet y . Funkcja podziału jest zdefiniowana według wzoru 3.

$$\text{split}(X, y) = \arg \max_{\text{cecha } c} \text{Gain}(X, y, c) \quad (3)$$

gdzie:

- c to cecha podziału,
- $\text{Gain}(X, y, c)$ to miara zysku informacyjnego dla danej cechy c .

Zysk informacyjny (Gain) zaprezentowany jako wzór 4 może być obliczony na podstawie entropii (miary niepewności) przed i po podziale.

$$\text{Gain}(X, y, c) = \text{Entropy}(y) - \sum_{v \in \text{Values}(c)} \frac{|X_v|}{|X|} \times \text{Entropy}(y_v) \quad (4)$$

gdzie:

- $\text{Entropy}(y)$ to entropia dla zbioru etykiet y ,
- $\text{Values}(c)$ to możliwe wartości cechy c ,
- X_v to podzbiór obiektów X dla których cecha c przyjmuje wartość v ,
- y_v to etykiety odpowiadające obiektom w X_v .

Podział ten jest realizowany w sposób binarny, co prowadzi do hierarchicznego drzewa decyzyjnego. Podczas klasyfikacji nowego obiektu, algorytm przechodzi przez drzewo, porównując cechy obiektu z warunkami podziału, aż do osiągnięcia liścia drzewa, który reprezentuje przewidywaną klasę.

Drzewa decyzyjne są wykorzystywane jako klasyfikatory binarne w węzłach struktury zagnieżdżonej dychotomii. Każde drzewo decyzyjne jest odpowiedzialne za rozróżnienie między dwiema podklasami danej klasy na podstawie określonej cechy.

Algorytm 1 Zagnieżdżona Dychotomia

```

1: function ZAGNIEŻDŻONA_DYCHOTOMIA( $X, y$ )
2:   if liczba_klas( $y$ )  $\leq 1$  then
3:     return liść z etykietą klasową  $y[0]$ 
4:   end if
5:   Wybierz najlepszą cechę podziału na podstawie analizy istotności cech
6:   Podziel zbiór obiektów  $X$  na dwie podgrupy na podstawie wybranej cechy
7:    $lewa\_X \leftarrow \{x \in X : \text{cecha}(x) \text{ spełnia pewien warunek}\}$ 
8:    $lewa\_y \leftarrow \{y_i : x_i \in lewa\_X\}$ 
9:    $prawa\_X \leftarrow \{x \in X : \text{cecha}(x) \text{ nie spełnia tego samego warunku}\}$ 
10:   $prawa\_y \leftarrow \{y_i : x_i \in prawa\_X\}$ 
11:  Utwórz węzeł decyzyjny dla wybranej cechy
12:   $lewy\_dziecko \leftarrow \text{ZagnieżdżonaDychotomia}(lewa\_X, lewa\_y)$ 
13:   $prawy\_dziecko \leftarrow \text{ZagnieżdżonaDychotomia}(prawa\_X, prawa\_y)$ 
14:  return węzeł decyzyjny z  $lewy\_dziecko$  i  $prawy\_dziecko$ 
15: end function

```

2.2 Plan eksperymentu

2.2.1 Założenia eksperymentów Zastosowanie eksperymentów pozwoli wyłonić najlepsze metody preprocessingu oraz dobrać hiperparametry drzew decyzyjnych. Wykorzystane wyniki z eksperymentów w zagnieżdżonej dychotomii pozwolą na uzyskanie najlepszego wyniku klasyfikacji.

2.2.2 Research questions Jak różne metody wyodrębniania cech obrazowych wpływają na skuteczność klasyfikacji owoców i warzyw? Celem badania jest zbadanie wpływu różnych metod wyodrębniania cech obrazowych na dokładność klasyfikacji owoców i warzyw. W szczególności zostaną przeanalizowane trzy różne techniki wyodrębniania cech obrazowych: VGG16, InceptionV3 i MobileNetV2. Badanie będzie polegać na porównaniu wyników klasyfikacji uzyskanych przy użyciu tych metod wyodrębniania cech, zastosowanych w modelu zagnieżdżonej dychotomii. Do oceny skuteczności klasyfikacji zostaną wykorzystane metryki takie jak dokładność, precyzja, czułość i miara F1. Dodatkowo, wyniki zostaną porównane za pomocą testów statystycznych, aby ocenić, która metoda wyodrębniania cech zapewnia najlepszą skuteczność klasyfikacji owoców i warzyw.

Czy zagnieżdżona dychotomia z usuwaniem klas sklasyfikowanych osiąga lepszą skuteczność niż zagnieżdżona dychotomia, w której wykorzystywane są wszystkie dane w każdym węźle?

Celem eksperymentu jest zbadanie, która implementacja zagnieżdżonej dychotomii osiąga lepsze wyniki. Zostaną sprawdzone dwie implementacje: zagnieżdżona dychotomia, gdzie każdy węzeł uczony jest tylko na danych wcześniej niesklasyfikowanych jako klasa znajdująca się w liściu oraz zagnieżdżona dychotomia, gdzie każdy węzeł jest uczony na wszystkich dostępnych danych z dopasowaniem klas. Jakość modeli będzie oceniana za pomocą miar takich jak

dokładność, precyzja, czułość oraz miara F1. Wyniki będą analizowane i wizualizowane, aby określić, która implementacja najlepiej poprawia jakość modelu.

Jak metody balansowania danych wpływają na jakość modelu?

W badaniu oceniony zostanie wpływ różnych metod balansowania danych na jakość modelu klasyfikacji obrazów owoców i warzyw. Przeanalizowane zostanie pięć metod balansowania danych: RandomOverSampler, SMOTE, RandomUnderSampler, TomekLinks oraz SMOTETomek. Cechy danych obrazowych zostaną wyodrębnione za pomocą metody wybranej z pierwszego eksperymentu. Następnie różne metody balansowania danych będą stosowane przed trenowaniem modelu zagnieżdżonej dychotomii z drzewami decyzyjnymi. Wydajność modelu będzie oceniana za pomocą miar takich jak dokładność, precyzja, czułość oraz miara F1. Wyniki będą analizowane i wizualizowane, aby określić, która metoda najlepiej poprawia jakość modelu.

Jak dobór parametrów wpływa na dokładność modelu?

Przeprowadzona zostanie seria eksperymentów, w których zostaną dobrane różne zestawy parametrów modelu, takie jak maksymalna głębokość drzewa decyzyjnego, minimalna liczba próbek wymagana do podziału węzła i maksymalna liczba cech branych pod uwagę podczas poszukiwania najlepszego podziału. Dobór parametrów będzie odbywał się za pomocą walidacji krzyżowej na zbiorze walidacyjnym, optymalizacja parametrów będzie wykonywana przez GridSearchCV, RandomizedSearchCV i BayesSearchCV, a dokładność modelu będzie monitorowana w trakcie procesu uczenia. W eksperymencie wykorzystywane będą metody przetwarzania danych i balansowania zbioru wybrane na podstawie pierwszego i drugiego eksperymentu.

2.2.3 Zestawy danych Wybrany zestaw danych został udostępniony na platformie Kaggle i nosi nazwę "Fruits and Vegetables Image Recognition Dataset". Składa się on z obrazów produktów spożywczych takich jak owoce i warzywa. Zestaw zawiera podział na 10 gatunków owoców oraz 26 gatunków warzyw co daje łącznie 36 klas produktów. Główną cechą, którą możemy wyróżnić w tym zbiorze danych jest kolor produktu. Dane są reprezentowane jako obrazy w formacie RGB, gdzie każdy piksel jest opisany trzema wartościami określającymi jego kolor. Analizowane będą także cechy kształtu produktów spożywczych. Kształt może być istotnym czynnikiem w procesie klasyfikacji, zwłaszcza w przypadku podobnych kolorów. Dodatkowo analizowane będą także cechy takie jak tekstura oraz rozmiar. Zbiór danych składa się z 3825 obrazów podzielonych na trzy kategorie takie jak: treningowa, która zawiera 3115 obrazów, testowa, która zawiera 359 obrazów oraz walidacyjna, która składająca się z 351 obrazów. Obrazy z tych trzech kategorii zostały scalone w jeden zbiór, ponieważ podział danych w modelu będzie odbywać się przy pomocy walidacji krzyżowej. Zbiór danych został utworzony za pomocą scrapowania danych z użyciem Bing Image Search. Większość obrazów zapisane jest w formacie JPG. Pozostałe w formacie PNG oraz JPEG.

2.2.4 Plan eksperymentu

Eksperyment 1: Wpływ metod wyodrębniania cech obrazowych na skuteczność klasyfikacji

W pierwszym eksperymencie zostanie zbadany wpływ różnych metod wyodrębniania cech obrazowych na skuteczność klasyfikacji owoców i warzyw. Najpierw zostanie przygotowany zbiór danych obrazów owoców i warzyw. Następnie za pomocą przetrenowanych modeli VGG16, InceptionV3 i MobileNetV2 zostaną wyodrębnione cechy z obrazów. Kolejnym krokiem będzie wykorzystanie modelu zagnieżdżonej dychotomii do klasyfikacji obrazów na podstawie wyodrębnionych cech. Skuteczność klasyfikacji zostanie oceniona za pomocą metryk takich jak dokładność, precyzja, czułość i miara F1. Wyniki uzyskane przy użyciu różnych metod wyodrębniania cech zostaną porównane za pomocą testów statystycznych, takich jak test t-studenta, aby ocenić, która metoda zapewnia najlepszą skuteczność klasyfikacji.

Eksperyment 2: Wpływ różnych implementacji zagnieżdżonej dychotomii na skuteczność klasyfikacji

W drugim eksperymencie zostanie zbadany wpływ różnych implementacji zagnieżdżonej dychotomii na skuteczność klasyfikacji. Wykorzystane zostaną cechy wyodrębnione za pomocą najlepiej sprawdzającej się metody z pierwszego eksperymentu. Zostaną przetestowane dwie implementacje zagnieżdżonej dychotomii: jedna, gdzie każdy węzeł jest uczony tylko na danych wcześniej niesklasyfikowanych jako klasa znajdująca się w liściu, oraz druga, gdzie każdy węzeł jest uczony na wszystkich dostępnych danych z dopasowaniem klas. Skuteczność obu implementacji zostanie oceniona za pomocą metryk takich jak dokładność, precyzja, czułość i miara F1. Wyniki zostaną porównane, aby określić, która implementacja najlepiej poprawia jakość modelu. Zostanie użyty test t-studenta, aby ocenić, która implementacja zapewnia najlepszą jakość klasyfikacji.

Eksperyment 3: Wpływ metod balansowania danych na jakość modelu

W trzecim eksperymencie oceniony zostanie wpływ różnych metod balansowania danych na jakość modelu klasyfikacji obrazów owoców i warzyw. Wykorzystane zostaną cechy wyodrębnione za pomocą najlepiej sprawdzającej się metody z pierwszego eksperymentu. Zostanie przetestowanych pięć metod balansowania danych: RandomOverSampler, SMOTE, RandomUnderSampler, TomekLinks oraz SMOTETomek. Model zagnieżdżonej dychotomii z drzewami decyzyjnymi zostanie przeszkolony na danych przetworzonych za pomocą tych metod. Skuteczność modelu zostanie oceniona za pomocą metryk takich jak dokładność, precyzja, czułość i miara F1. Końcowo zostanie użyty test t-studenta w celu sprawdzenia różnic statystycznych. Wyniki zostaną porównane, aby określić, która metoda balansowania danych najlepiej poprawia jakość modelu.

Eksperyment 4: Wpływ doboru parametrów na dokładność modelu

W czwartym eksperymencie zostanie przeprowadzona seria eksperymentów, w których zostaną dobrane różne zestawy parametrów modelu, takie jak maksymalna głębokość drzewa decyzyjnego, minimalna liczba próbek wymagana do podziału węzła i maksymalna liczba cech branych pod uwagę podczas poszukiwania najlepszego podziału. Dobór parametrów będzie odbywał

się za pomocą walidacji krzyżowej na zbiorze walidacyjnym, a optymalizacja parametrów będzie wykonywana za pomocą GridSearchCV, RandomizedSearchCV i BayesSearchCV. Skuteczność modelu będzie monitorowana w trakcie procesu uczenia za pomocą metryk takich jak dokładność, precyzja, czułość i miara F1. Końcowo zostanie użyty test t-studenta w celu sprawdzenia różnic statystycznych. W eksperymencie wykorzystywane będą metody przetwarzania danych i balansowania zbioru wybrane na podstawie wyników z poprzednich eksperymentów.

2.2.5 Środowisko eksperymentalne Do przeprowadzenia eksperymentów wykorzystany zostanie komputer wyposażony w procesor AMD Ryzen 7 5800X oraz 32 GB pamięci RAM. Systemem operacyjnym użytym w eksperymentach będzie 64-bitowy system, z procesorem obsługującym architekturę x64. Dodatkowo w ramach potrzeby zastosowane zostanie środowisko Google Colab umożliwiające uruchomienie kodu w chmurze obliczeniowej. Eksperyment zostanie wykonany w języku Python w wersji 3.11, wykorzystana zostanie biblioteka do uczenia maszynowego scikit-learn 1.4.1 oraz narzędzia do przetwarzania danych takie jak pandas 2.2.1 i numpy.

3 Przedstawienie wyników badań

3.1 Wyniki badań i analizy statystycznej

3.1.1 Wpływ metod wyodrębniania cech obrazowych Po przeprowadzeniu eksperymentów i analizie wyników, uzyskano następujące rezultaty przedstawione w tabeli 1.

Table 1. Wyniki klasyfikacji dla różnych metod wyodrębniania cech.

Metoda	Accuracy	Macro Avg Precision	Macro Avg Recall	Macro Avg F1 Score
VGG16	0.53	0.654	0.53	0.566
InceptionV3	0.364	0.512	0.362	0.404
MobileNetV2	0.386	0.518	0.386	0.422

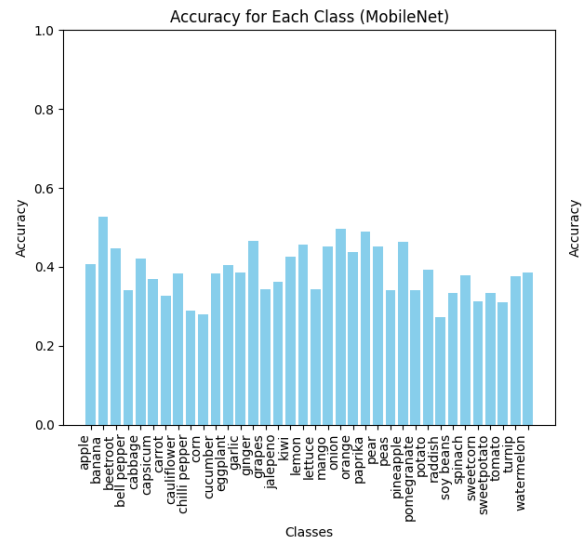


Fig. 2. Wykresy przedstawiające średnią dokładność dla MobileNet z podziałem na klasy.

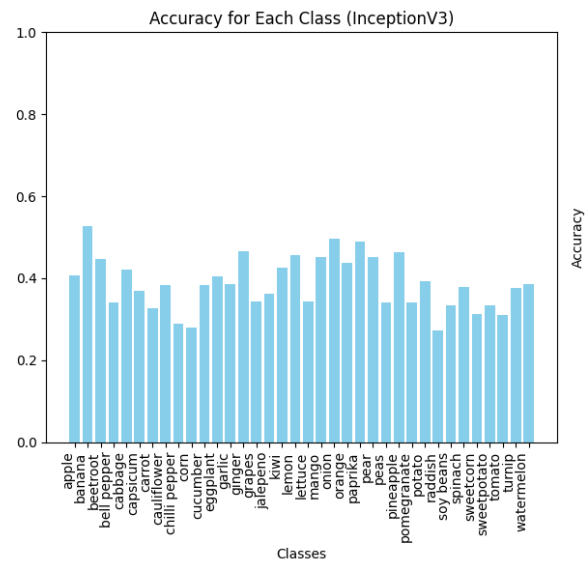


Fig. 3. Wykresy przedstawiające średnią dokładność dla InceptionV3 z podziałem na klasy.

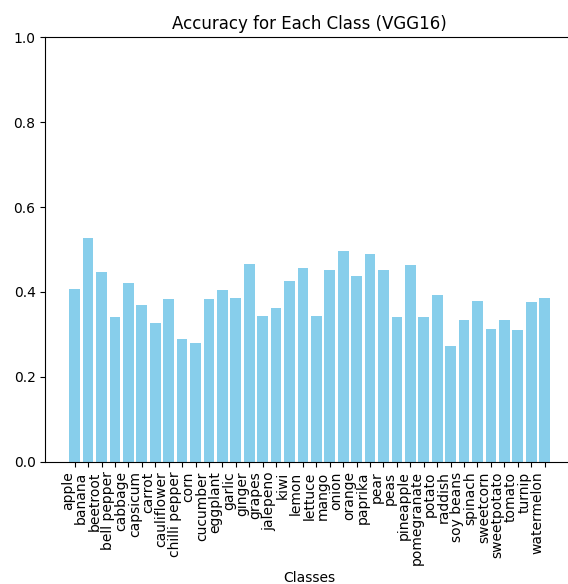


Fig. 4. Wykresy przedstawiające średnią dokładność dla VGG16 z podziałem na klasy.

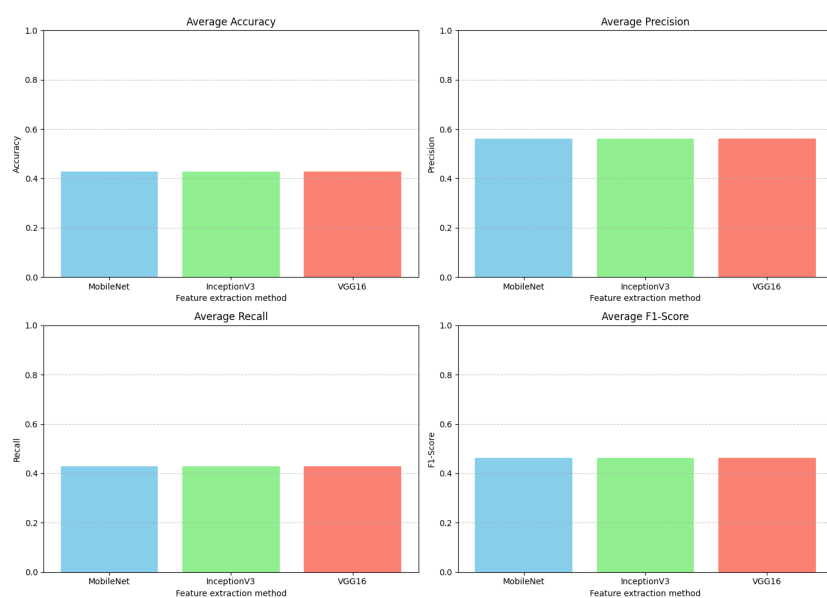
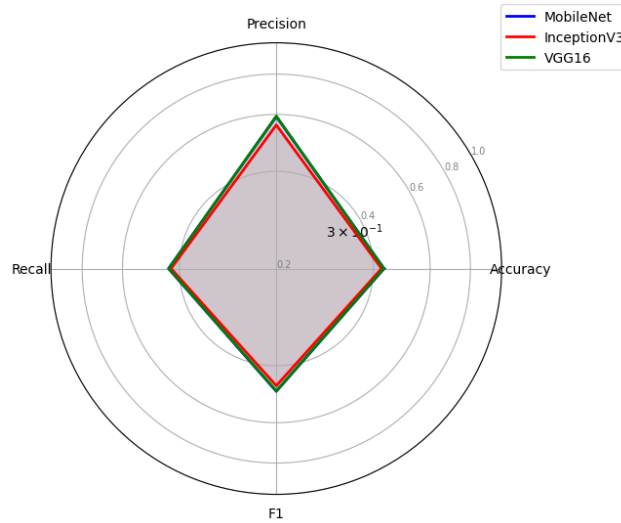


Fig. 5. Wykres przedstawiający średnie wyniki dla każdej metody.

Wykres radarowy dla Accuracy, Precision, Recall, F1

**Fig. 6.** Radar Chart przedstawiający średnie wyniki dla każdej metody.**Table 2.** Porównanie statystyk dla par metod ekstrakcji cech

Porównanie	T-statystyka	P-wartość
MobileNet vs. InceptionV3	0.6693	0.5519
MobileNet vs. VGG16	0.018	0.987
InceptionV3 vs. VGG16	-0.64	0.57

3.1.2 Wpływ różnych implementacji zagnieżdżonej dychotomii na skuteczność klasyfikacji Po przeprowadzeniu eksperymentów i analizy wyników, uzyskano następujące rezultaty przedstawione w tabeli 3.

Table 3. Wyniki klasyfikacji dla różnych metod balansowania danych.

Metoda wyodrębniania cech	Accuracy	Precision	Recall	F1-Score
With Removing	0.476	0.515	0.474	0.480
Without Removing	0.472	0.625	0.472	0.516

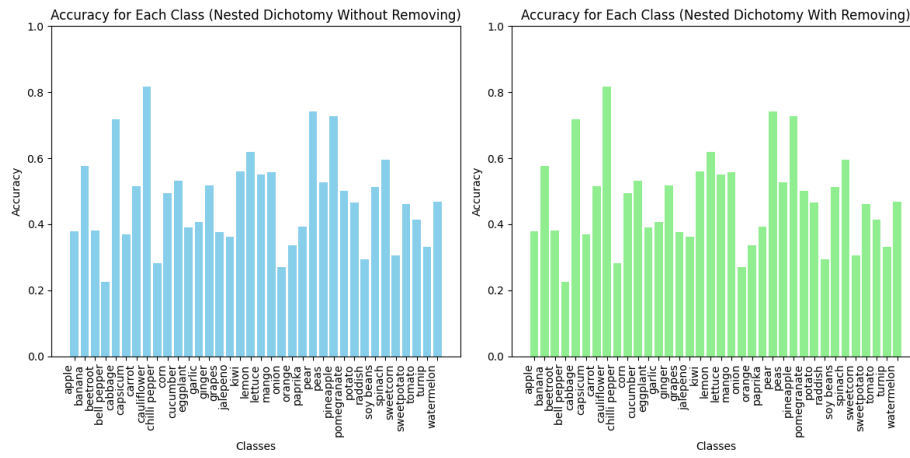


Fig. 7. Wykresy przedstawiające średnią dokładność dla każdej metody z podziałem na klasy.

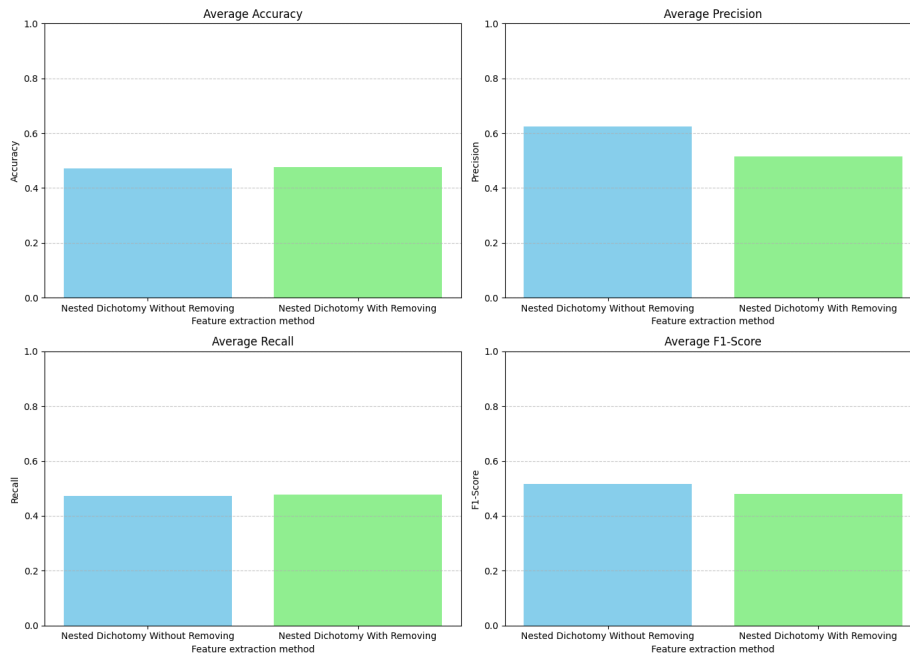


Fig. 8. Wykres przedstawiający średnie wyniki dla każdej metody.

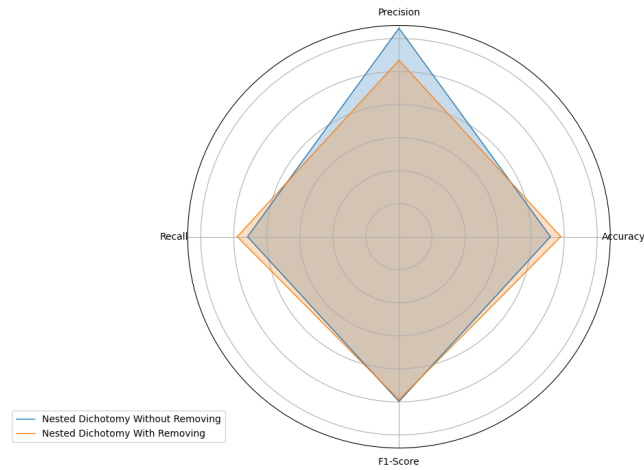


Fig. 9. Radar Chart przedstawiający średnie wyniki dla każdej metody.

Table 4. Porównanie wyników testów statystycznych dla implementacji zagnieżdżonej dychotomii

Porównanie	T-statystyka	P-wartość
Without Removing vs. With Removing	-1.692	0.34

3.1.3 Wpływ metod balansowania danych Po przeprowadzeniu eksperymentów i analizy wyników, uzyskano następujące rezultaty przedstawione w tabeli 5.

Table 5. Wyniki klasyfikacji dla różnych metod balansowania danych.

Metoda wyodrębniania cech	Accuracy	Precision	Recall	F1-Score
RandomOverSampler	0.578	0.717	0.578	0.620
SMOTE	0.560	0.657	0.560	0.584
RandomUnderSampler	0.152	0.175	0.152	0.106
TomekLinks	0.535	0.668	0.535	0.575
SMOTETomek	0.566	0.660	0.566	0.589

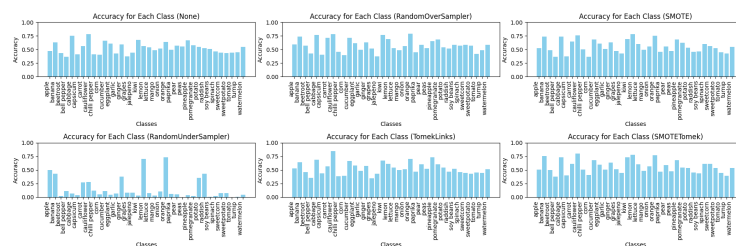


Fig. 10. Wykresy przedstawiające średnią dokładność dla każdej metody z podziałem na klasy.

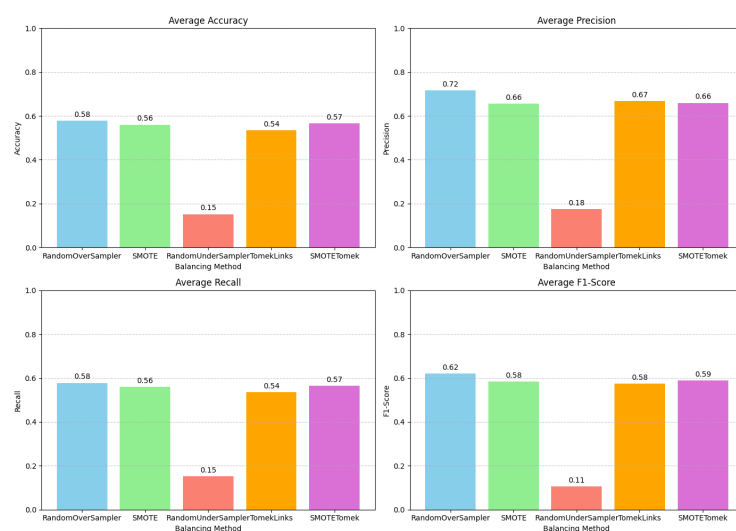


Fig. 11. Wykres przedstawiający średnie wyniki dla każdej metody.

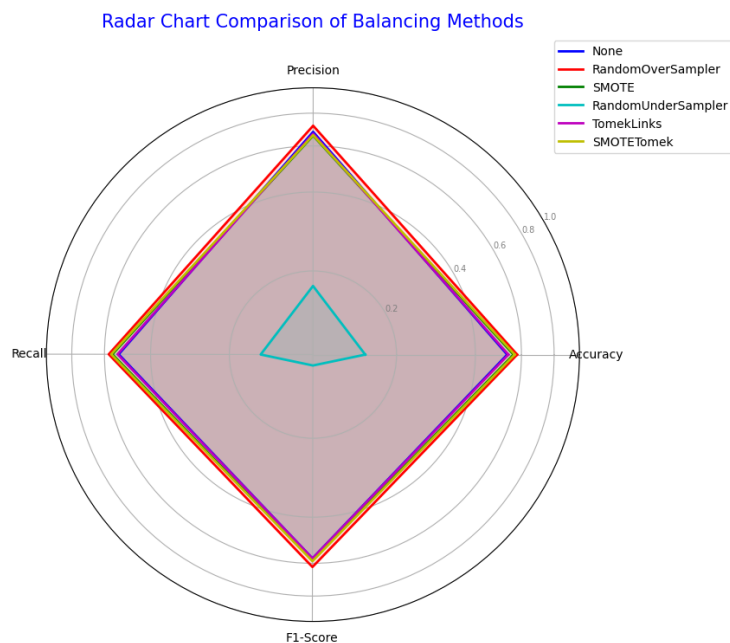


Fig. 12. Radar Chart przedstawiający średnie wyniki dla każdej metody.

Table 6. Porównanie statystyk dla każdej metody balansowania danych

Porównanie	T-statystyka	P-wartość
RandomOverSampler vs. SMOTE	4.636	0.010
RandomOverSampler vs. RandomUnderSampler	116.865	0.000
RandomOverSampler vs. TomekLinks	15.062	0.000
RandomOverSampler vs. SMOTETomek	3.645	0.022
SMOTE vs. RandomUnderSampler	99.561	0.000
SMOTE vs. TomekLinks	7.312	0.002
SMOTE vs. SMOTETomek	-1.514	0.205
RandomUnderSampler vs. TomekLinks	-121.872	0.000
RandomUnderSampler vs. SMOTETomek	-115.113	0.000
TomekLinks vs. SMOTETomek	-11.038	0.000

3.1.4 Wpływ doboru parametrów na dokładność modelu Po przeprowadzeniu eksperymentów i analizy wyników, uzyskano następujące rezultaty przedstawione w tabeli 7.

Table 7. Wyniki klasyfikacji dla różnych metod doboru parametrów.

	Accuracy	Macro Avg Precision	Macro Avg Recall	Macro Avg F1
GridSearchCV	0.574	0.717	0.574	0.619
RandomizedSearchCV	0.582	0.725	0.582	0.626
BayesSearchCV	0.569	0.715	0.569	0.616

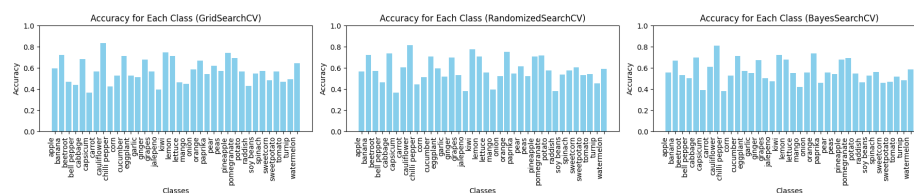


Fig. 13. Wykresy przedstawiające średnią dokładność dla każdej metody z podziałem na klasy.

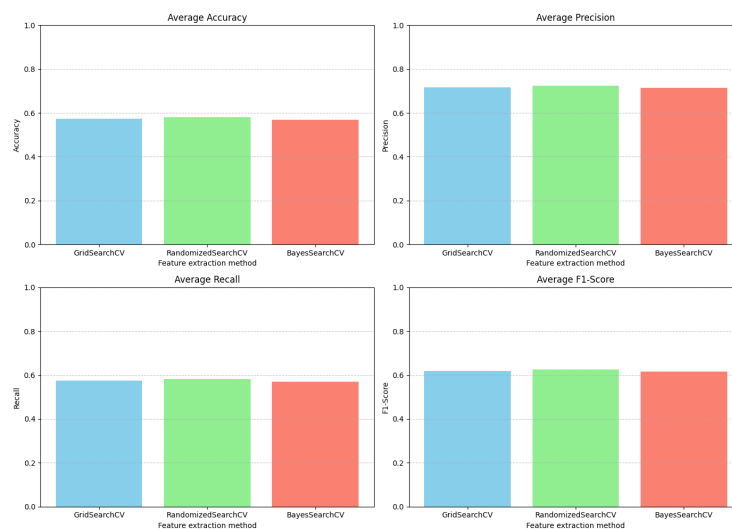


Fig. 14. Wykres przedstawiający średnie wyniki dla każdej metody.

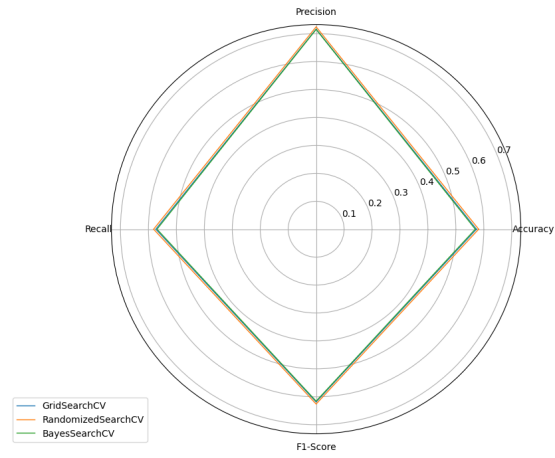


Fig. 15. Radar Chart przedstawiający średnie wyniki dla każdej metody.

Table 8. Najczęściej wybierane wartości parametrów dla poszczególnych metod

Metoda	Parametr	Najczęstsza wartość
GridSearchCV	criterion	entropy
	max_depth	None
	min_samples_leaf	1
	min_samples_split	2
	splitter	random
RandomizedSearchCV	criterion	log_loss
	max_depth	30
	min_samples_leaf	2
	min_samples_split	10
	splitter	best
BayesSearchCV	criterion	entropy
	max_depth	30
	min_samples_leaf	1
	min_samples_split	15
	splitter	best

Table 9. Porównanie wyników testów statystycznych dla różnych metod doboru hiperparametrów

Porównanie	T-statystyka	P-wartość
GridSearchCV vs. RandomizedSearchCV	-1.108	0.33
GridSearchCV vs. BayesSearchCV	0.654	0.549
RandomizedSearchCV vs. BayesSearchCV	1.64	0.176

3.2 Analiza wyników i wnioski

3.2.1 Wpływ metod wyodrębniania cech obrazowych Badanie miało na celu porównanie skuteczności trzech różnych metod wyodrębniania cech obrazowych (VGG16, InceptionV3 i MobileNetV2) w klasyfikacji owoców i warzyw. W celu oceny skuteczności klasyfikacji wykorzystano metryki takie jak dokładność, precyzja, czułość i miara F1.

Z analizy wyników przedstawionych w tabeli 1 oraz na wykresach 2 i 5 wynika, że model oparty na architekturze VGG16 osiągnął najwyższą dokładność klasyfikacji oraz najlepsze wyniki w większości badanych metryk. Jednakże, testy statystyczne przedstawione w tabeli 2 wykazały, że różnice między metodami nie są istotne statystycznie, co sugeruje, że żadna z badanych metod wyodrębniania cech nie zapewnia zdecydowanej przewagi w kontekście klasyfikacji owoców i warzyw.

Wizualizacja wyników za pomocą radar chart 6 pozwala na szybkie porównanie średnich wyników dla każdej metody, co potwierdza przewagę VGG16, ale także pokazuje, że MobileNetV2 osiąga podobne wyniki do InceptionV3.

Wnioskiem z przeprowadzonych badań jest fakt, że wybór metody wyodrębniania cech obrazowych może mieć wpływ na skuteczność klasyfikacji, ale istotność tej różnicy może być ograniczona w przypadku konkretnego problemu klasyfikacyjnego, takiego jak klasyfikacja owoców i warzyw. Dalsze badania mogą być konieczne w celu potwierdzenia tych wyników na większych zbiorach danych oraz dla innych klas problemów klasyfikacyjnych.

3.2.2 Wpływ różnych implementacji zagnieżdżonej dychotomii na skuteczność klasyfikacji Analiza wyników eksperymentu 3 sugeruje, że metoda zagnieżdżonej dychotomii bez usuwania klas sklasyfikowanych może być bardziej efektywna w kontekście klasyfikacji. Pomimo nieco niższej średniej dokładności w porównaniu do metody z usuwaniem klas, ta pierwsza metoda osiągnęła istotnie wyższą precyzję (0.631 w porównaniu do 0.534) oraz podobny poziom miary F1 (0.499 w porównaniu do 0.494).

Wysoka precyzja sugeruje, że metoda bez usuwania klas lepiej radzi sobie z minimalizacją fałszywie pozytywnych przypadków, co może być kluczowe w wielu zastosowaniach, szczególnie gdy istotne są konsekwencje wynikające z błędnej klasyfikacji pozytywnych przypadków. Ponadto, podobny poziom miary F1 dla obu metod wskazuje na równowagę między zdolnością modelu do poprawnej identyfikacji pozytywnych przypadków a minimalizacją liczby fałszywie pozytywnych przypadków, co jest istotnym czynnikiem w ocenie jakości klasyfikacji.

Wizualizacja wyników za pomocą radar chart 9 pozwala na szybkie porównanie średnich wyników dla obu metod, co potwierdza przewagę metody bez usuwania.

Ostatecznie, pomimo nieznacznie niższej dokładności, metoda bez usuwania klas sklasyfikowanych może być preferowana ze względu na jej lepszą precyzję oraz zbliżoną do metody z usuwaniem klas wartość miary F1. Jednakże, zawsze należy uwzględniać kontekst danego problemu oraz indywidualne wymagania i preferencje dotyczące jakości klasyfikacji.

3.2.3 Wpływ metod balansowania danych na jakość modelu Analiza wyników przedstawionych w tabeli 5 oraz na wykresach 10 i 11 oraz za pomocą radar chart 12 wskazuje, że metoda RandomOverSampler osiągnęła najwyższe wartości we wszystkich miarach (dokładność, precyzja, czułość i F1-Score), co sugeruje, że jest najbardziej efektywną metodą balansowania danych dla klasyfikacji obrazów owoców i warzyw. Zastosowanie tej metody poprawiło jakość modelu w porównaniu do innych metod balansowania danych, co zostało potwierdzone przez istotne statystycznie różnice w większości porównań.

Metoda SMOTETomek również wykazała wysoką efektywność, z wynikami zbliżonymi do RandomOverSampler, jednak brak istotnych statystycznie różnic 6 między tymi dwiema metodami sugeruje, że obie mogą być równie skuteczne.

Z kolei metoda RandomUnderSampler uzyskała najniższe wyniki we wszystkich miarach, co sugeruje, że nie jest skuteczną metodą balansowania danych w tym kontekście. Metoda TomekLinks osiągnęła wyniki lepsze niż RandomUnderSampler, ale gorsze niż RandomOverSampler, SMOTE i SMOTETomek, co sugeruje jej umiarkowaną skuteczność.

Podsumowując, wybór odpowiedniej metody balansowania danych ma istotny wpływ na jakość modelu klasyfikacji obrazów. W kontekście analizowanego problemu, RandomOverSampler i SMOTETomek okazały się być najbardziej efektywnymi metodami, podczas gdy RandomUnderSampler okazał się być najmniej skuteczny.

3.2.4 Wpływ doboru parametrów na dokładność modelu Analiza wyników przeprowadzonych eksperymentów przedstawionych w tabeli 7 oraz na wykresach 13 i 14 oraz za pomocą radar chart 12 wskazuje, że wybór metody optymalizacji hiperparametrów nie ma istotnego wpływu na jakość modelu klasyfikacji. Pomimo niewielkich różnic w średnich miarach jakości pomiędzy badanymi metodami, żadna z nich nie wykazała znaczącej przewagi nad pozostałymi. Także przeprowadzane testy statystyczne 9 nie wykazały istotnych różnic pomiędzy metodami.

Wnioskiem z przeprowadzonego badania jest stwierdzenie, że w przypadku analizowanej klasyfikacji, dobór konkretnego algorytmu optymalizacji hiperparametrów może być mniej istotny niż inne czynniki, takie jak wybór właściwego modelu czy przetwarzanie danych wejściowych. Najlepsze zestawy parametrów, wybrane na podstawie częstości ich występowania wśród zestawów

parametrów wybieranych jako optymalne w trakcie procesu optymalizacji hiperparametrów, dla poszczególnych metod przedstawiono w tabeli 8.

Analiza częstości występowania najlepszych parametrów dla każdej z trzech metod optymalizacji hiperparametrów (GridSearchCV, RandomizedSearchCV i BayesSearchCV) wskazuje na pewne ogólne tendencje. Najlepsze parametry, uwzględniając wszystkie metody, obejmują kryterium podziału jako entropię, maksymalną głębokość drzewa równą 30, minimalną liczbę próbek na liściu wynoszącą 1, minimalną liczbę próbek do podziału równą 2, oraz metodę podziału węzła wybieraną jako best. Te wyniki sugerują, że te parametry mogą być bardziej uniwersalne i skuteczne w kontekście różnych metod optymalizacji hiperparametrów dla modeli klasyfikacji opartych na drzewach decyzyjnych.

Należy jednak zauważyć, że przeprowadzone eksperymenty obejmowały tylko ograniczoną liczbę zestawów danych i modeli. Dalsze badania, obejmujące większą różnorodność zbiorów danych oraz różne rodzaje modeli, mogą dostarczyć bardziej szczegółowych wniosków na temat wpływu różnych metod optymalizacji hiperparametrów na różne warunki zadania klasyfikacji.

3.3 Wnioski końcowe

Przeprowadzone badania nad klasyfikacją obrazów owoców i warzyw pozwoliły na wyciągnięcie kilku istotnych wniosków. Przede wszystkim, wybór metod wyodrębniania cech obrazowych, takich jak VGG16, InceptionV3 i MobileNetV2, wpływa na skuteczność klasyfikacji. Chociaż VGG16 osiągnęła najlepsze wyniki, różnice pomiędzy metodami nie były istotne statystycznie, co sugeruje, że nie ma jednej metody wyraźnie lepszej od innych w tym kontekście.

Dodatkowo, analiza implementacji zagnieżdżonej dychotomii wykazała, że metoda bez usuwania klas sklasyfikowanych może oferować wyższą precyzję, co jest istotne w minimalizacji fałszywie pozytywnych klasyfikacji. Mimo nieznacznie niższej dokładności, metoda ta może być preferowana ze względu na lepszą precyzję.

W zakresie balansowania danych, RandomOverSampler i SMOTETomek okazały się najbardziej efektywne, poprawiając jakość modelu w porównaniu do innych metod. Natomiast RandomUnderSampler był najmniej skuteczny. Wybór odpowiedniej metody balansowania danych jest kluczowy dla osiągnięcia wysokiej jakości modelu klasyfikacji.

Na koniec, wybór metody optymalizacji hiperparametrów nie miał istotnego wpływu na jakość modelu klasyfikacji. Sugeruje to, że w analizowanym przypadku ważniejsze mogą być inne czynniki, takie jak wybór modelu i przetwarzanie danych wejściowych.

3.4 Możliwości rozwoju

Wyniki niniejszego badania wskazują na kilka kierunków, które mogą stanowić podstawę dla przyszłych prac i rozwoju w obszarze klasyfikacji obrazów. Przede wszystkim, przeprowadzenie badań na większych i bardziej zróżnicowanych

zbiorach danych może dostarczyć bardziej szczegółowych informacji na temat skuteczności poszczególnych metod i ich zastosowania w różnych kontekstach. Testowanie nowoczesnych architektur sieci neuronowych, takich jak EfficientNet, ResNet czy nasilenie badań nad modelami transformatorowymi, może prowadzić do odkrycia bardziej efektywnych metod wyodrębniania cech obrazowych.

Dalsza analiza metod optymalizacji hiperparametrów, obejmująca bardziej zaawansowane techniki, może przyczynić się do poprawy jakości modeli klasyfikacyjnych. Badanie kombinacji różnych technik balansowania danych lub rozwijanie nowych metod może prowadzić do dalszej poprawy jakości modeli w kontekście nie zrównoważonych zbiorów danych. Badanie wpływu specyficznych warunków aplikacyjnych, takich jak różne typy szumów w obrazach czy zmienne oświetlenie, może prowadzić do tworzenia bardziej robustnych modeli klasyfikacyjnych. Testowanie i wdrażanie opracowanych metod w rzeczywistych systemach rozpoznawania obrazów, takich jak systemy kontroli jakości w przemyśle spożywczym, może dostarczyć cennych informacji zwrotnych i przyczynić się do dalszego doskonalenia metod.

Dalsze badania w tych obszarach mogą znacząco przyczynić się do rozwoju bardziej skutecznych i wszechstronnych systemów klasyfikacji obrazów, co znajdzie zastosowanie w wielu dziedzinach przemysłu i nauki.

References

1. Koszela, K., Weres, J.: ANALIZA I KLASYFIKACJA OBRAZÓW SUSZU WARZYWNEGO Z WYKORZYSTANIEM SZTUCZNYCH SIECI NEURONOWYCH. Inżynieria Rolnicza 2/2005, Akademia Ekonomiczna w Krakowie, Instytut Inżynierii Rolniczej Akademia Rolnicza w Poznaniu. <https://bibliotekanauki.pl/articles/287257.pdf>.
2. Lemaître, G., Nogueira, F., Aridas, C. K.: SMOTETomek — Version 0.12.3 - imbalanced-learn. <https://imbalanced-learn.org/stable/references/generated/imblearn.combine.SMOTETomek.html>.
3. Brownlee, J.: How to Combine Oversampling and Undersampling for Imbalanced Classification. MachineLearningMastery. <https://machinelearningmastery.com/combine-oversampling-and-undersampling-for-imbalanced-classification/>.
4. Lemaître, G., Nogueira, F., Aridas, C. K.: Combination of over- and under-sampling — Version 0.12.3. imbalanced-learn. <https://imbalanced-learn.org/stable/combine.html>.
5. Appaji, N.: Balancing Act: Mastering Imbalanced Data with SMOTE and Tomek-Link Strategies. Medium. <https://medium.com/@niranjan.appaji/balancing-act-mastering-imbalanced-data-with-smote-and-tomek-link-strategies-289f39597122>.
6. Pedregosa, F. et al.: Tuning the hyper-parameters of an estimator - scikit-learn. https://scikit-learn.org/stable/modules/grid_search.html.
7. Gupta, L.: Comparison of Hyperparameter Tuning algorithms: Grid search, Random search, Bayesian optimization. Analytics Vidhya. <https://medium.com/analytics-vidhya/comparison-of-hyperparameter-tuning-algorithms-grid-search-random-search-bayesian-optimization-532>.
8. Noble, R.: Hyperparameter Tuning with Keras and GridSearchCV: A Comprehensive Guide. Zegami Tech. <https://medium.com/@AIandInsights/hyperparameter-tuning-with-keras-and-gridsearchcv-a-comprehensive-guide-46214cc0d999>.

9. Stack Overflow community: python 3.x - Gridsearchcv vs Bayesian optimization. Stack Overflow. <https://stackoverflow.com/questions/55849512/gridsearchcv-vs-bayesian-optimization>.
10. Cross Validated community: Hyper parameters tuning: Random search vs Bayesian optimization. Cross Validated. <https://stats.stackexchange.com/questions/302891/hyper-parameters-tuning-random-search-vs-bayesian-optimization>.
11. Khan, M. A., et al.: Multiclass skin lesion classification using deep learning networks optimal information fusion. Discover Applied Sciences. <https://oaji.net/articles/2022/9778-1657798324.pdf>.
12. Grzyb, M.: Wybór odpowiedniego algorytmu. Część 2 – algorytmy klasyfikacyjne. <https://mateuszgrzyb.pl/wybor-odpowiedniego-algorytmu-czesc-2-algorytmy-klasyfikacyjne/>.
13. Raschka, S., Mirjalili, V.: Python machine learning i deep learning: biblioteki scikit-learn i TensorFlow 2. https://www.biblos.pk.edu.pl/ST/2021/04/100000331444/100000331444_Raschka_PythonMachine.pdf.
14. Koszela, K., Weres, J.: Analiza i klasyfikacja obrazów suszu warzywnego z wykorzystaniem sztucznych sieci neuronowych. Inżynieria Rolnicza. <https://bibliotekanauki.pl/articles/287257>.
15. Koszela, K., Weres, J.: Neuronowa klasyfikacja obrazów suszu warzywnego. Inżynieria Rolnicza. <https://bibliotekanauki.pl/articles/287040.pdf>.
16. Weres, J.: Analiza I Klasyfikacja Obrazów Suszu Warzywnego Z Wykorzystaniem Sztucznych Sieci Neuronowych. Academia.edu. https://www.academia.edu/86628513/Analiza_I_Klasyfikacja_Obraz%C3%B3w_Suszu_Warzywnego_Z_Wykorzystaniem_Sztucznych_Sieci_Neuronowych.
17. Noble, R.: Comparing pre-trained deep learning models for feature extraction. Zegami Tech. <https://medium.com/zegami-tech/comparing-pre-trained-deep-learning-models-for-feature-extraction-c617da54641>.
18. Michalska, M.: MULTICLASS SKIN LESIONS CLASSIFICATION BASED ON DEEP NEURAL NETWORKS. Lublin University of Technology, Department of Electronics and Information Technology, Poland. <https://doi.org/10.35784/iapgos.2963>.