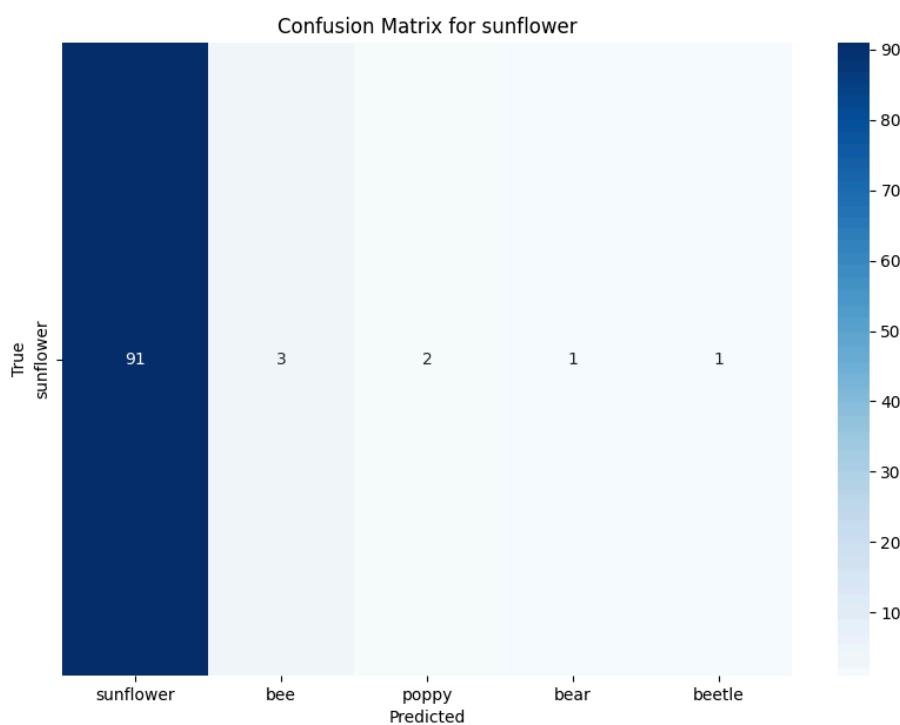
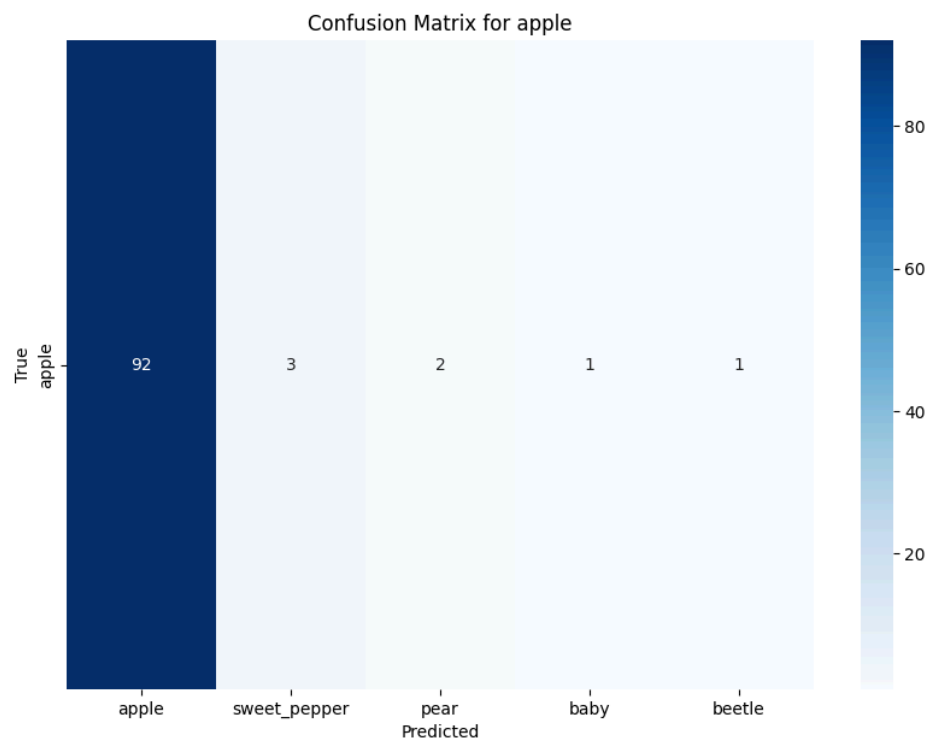


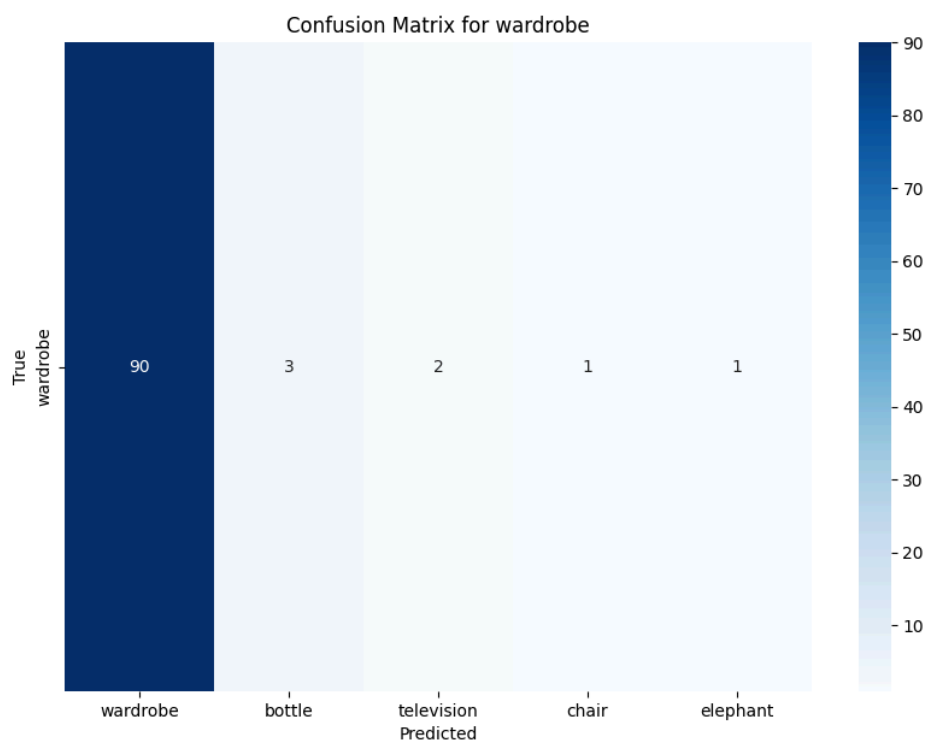
Najlepszy model - 66,30% dokładności predykcji

Macierze pomyłek

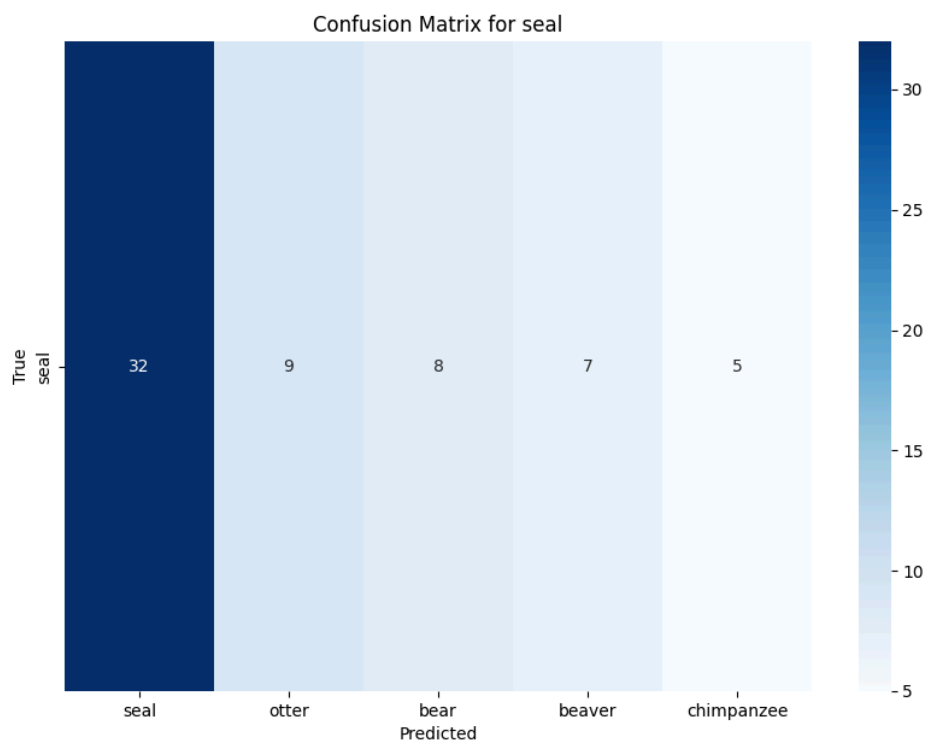
Przedstawione są macierze pomyłek dla trzech najlepiej przewidywanych i trzech najgorzej przewidywanych klas. Macierze pokazują 5 najczęściej przewidywanych klas przez model dla danej klasy, czyli także FN.

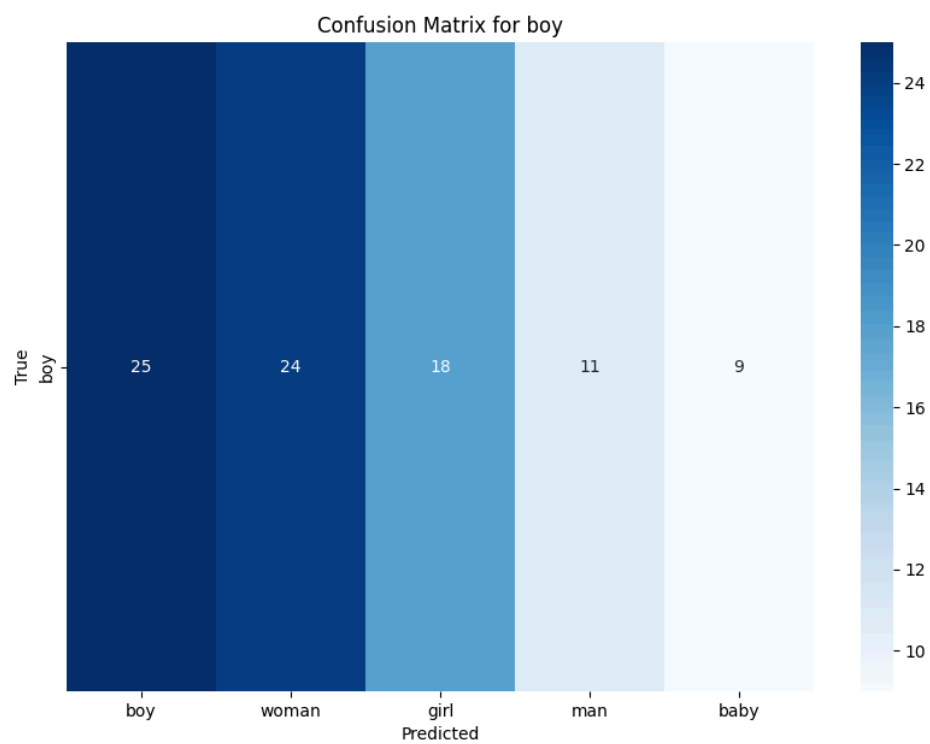
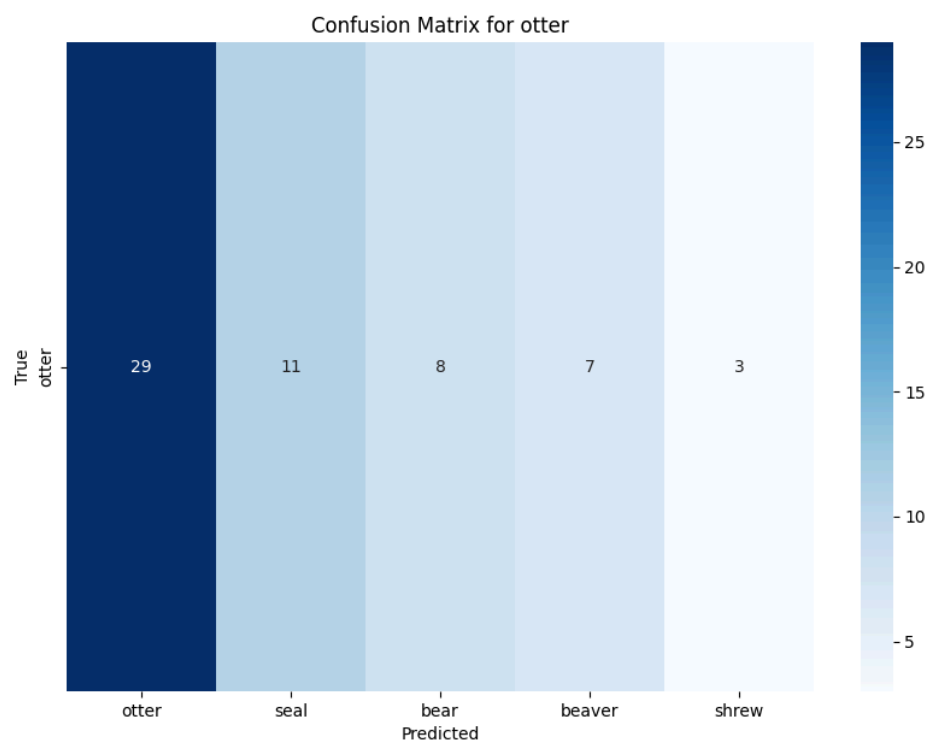
Najlepiej przewidywane klasy: **Apple (92%), Sunflower(91%), Wardrobe(90%)**





Najgorzej przewidywane klasy: **Seal(32%), Otter(29%), Boy(25%)**





Metryki oceny

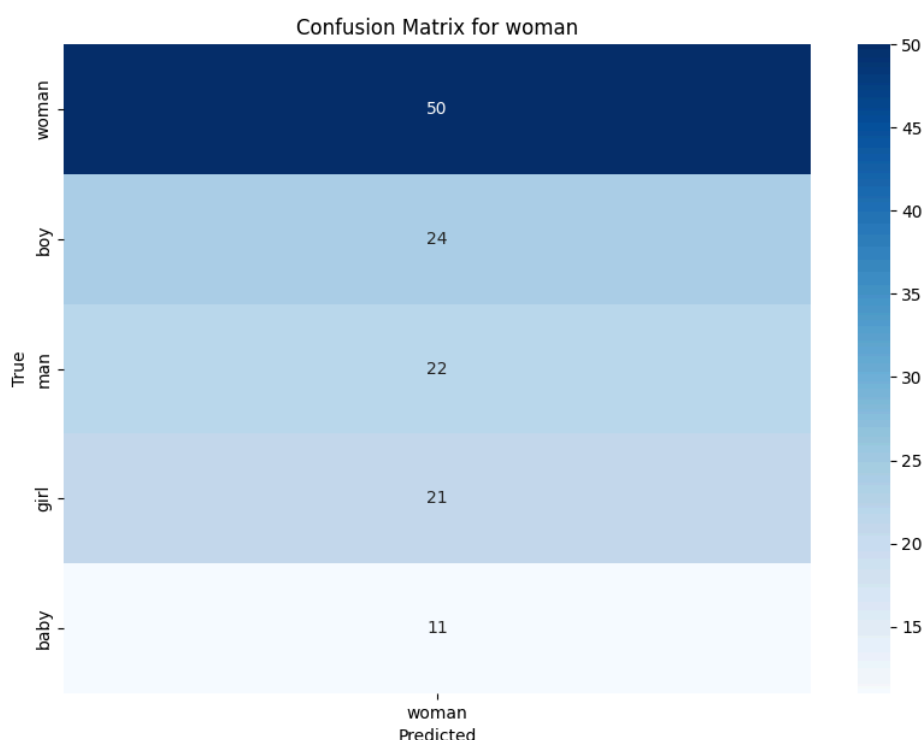
Precyzja (ang. Precision Score) - 0,675

$$Precision = \frac{TP}{TP+FP}$$

Określa ile z przewidzianych jako pozytywne przypadków rzeczywiście jest pozytywnych. Do obliczenia precyzji używałem wariantu ważonego, czyli precyzja była obliczana dla każdej klasy i ważona odpowiednio przez liczbę przykładów danej klasy w całym zbiorze.

Klasa z najlepszą precyzją: Trout (0,94)

Klasa z najmniejszą precyzją: Woman (0,34) - na wykresie poniżej można zobaczyć, że model przewiduje często fałszywie pozytywnie klasę woman:



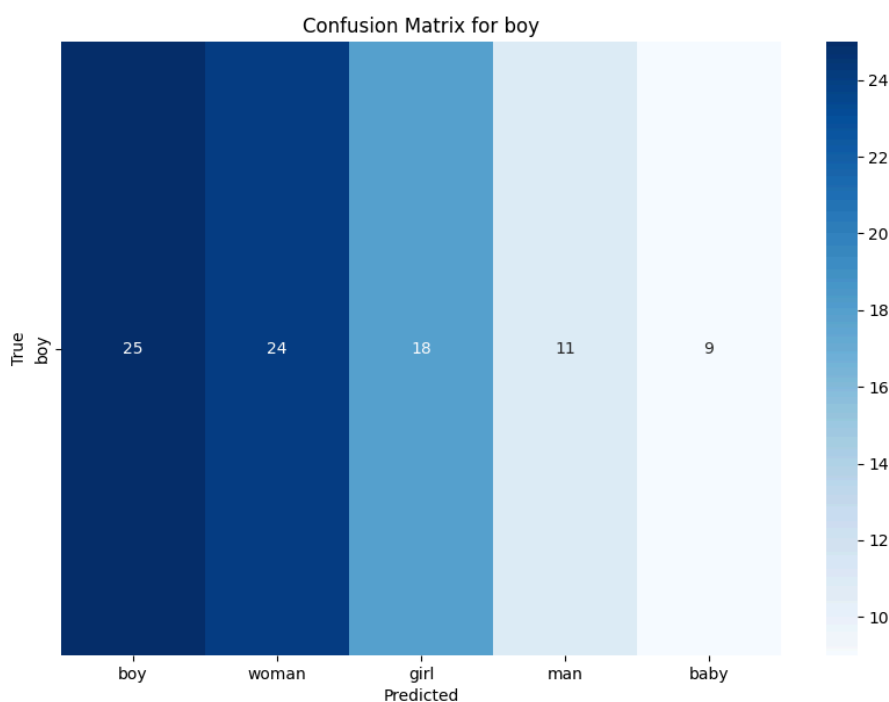
Czułość (ang. Recall) - 0,663

$$Recall = \frac{TP}{TP+FN}$$

Mierzy jaką część wszystkich rzeczywiście pozytywnych przypadków model wykrył. Jest to kluczowa metryka, gdy pominięcie pozytywnego przypadku może mieć poważne konsekwencje. Do policzenia czułości też używałem wariantu ważonego.

Klasa z najlepszą czułością: Apple (0,92)

Klasy z najmniejszą czułością: Boy (0,26) - jak widać na wykresie poniżej dla klasy Boy model przewiduje dużo fałszywie negatywnych(FN) klas:



F1-Score - 0,665

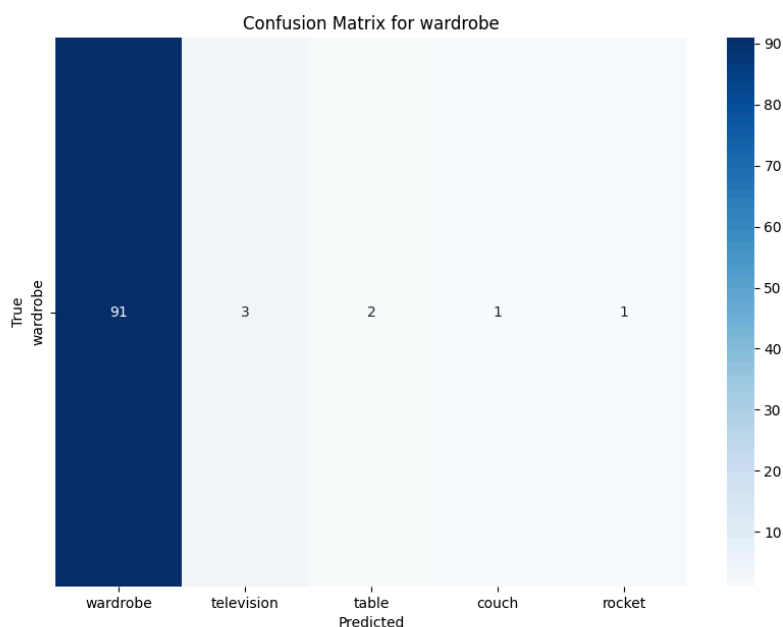
To średnia harmoniczna precyzji i czułości, która uwzględnia oba te aspekty. Do jej obliczenia używałem również wariantu ważonego. F1-score jest przydatny, gdy zależy nam na równowadze między precyzją a recall, np. w sytuacjach, gdzie zarówno false positives, jak i false negatives są kosztowne.

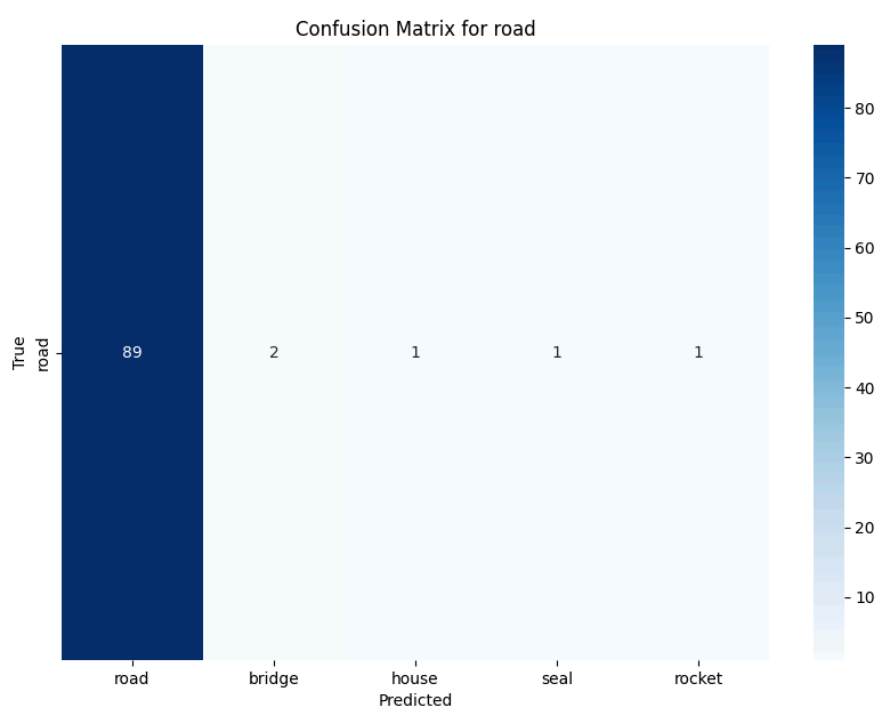
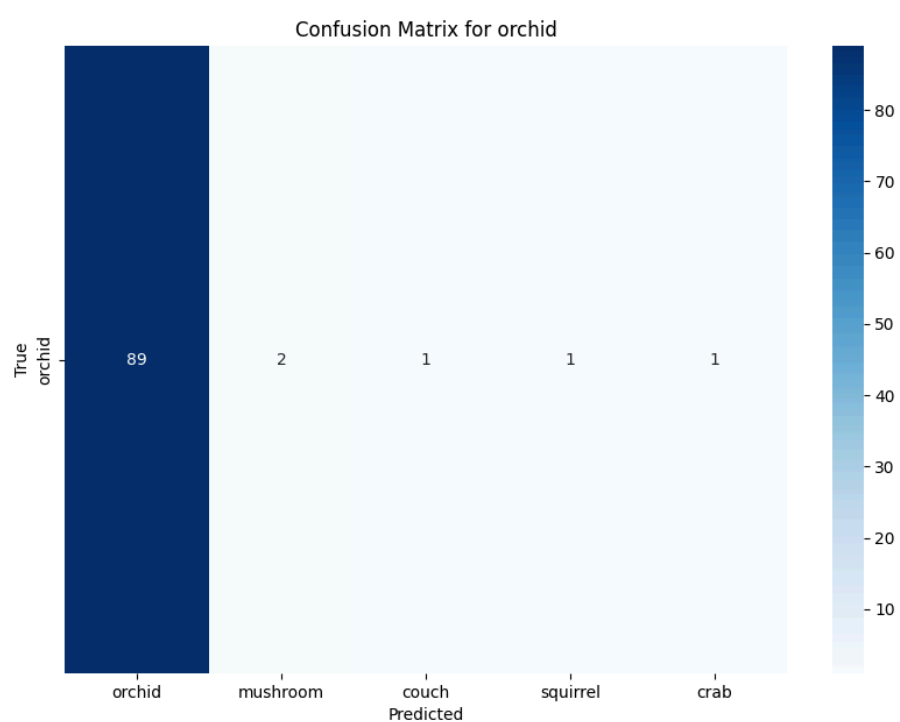
Drugi najlepszy model - 64,32% dokładności predykcji

Macierze pomyłek

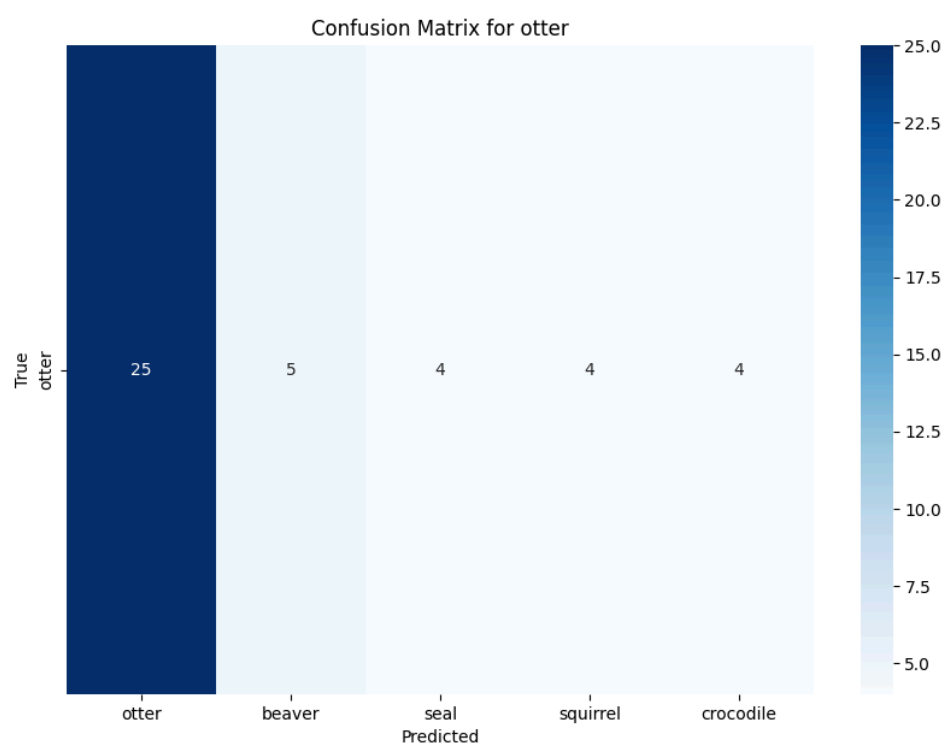
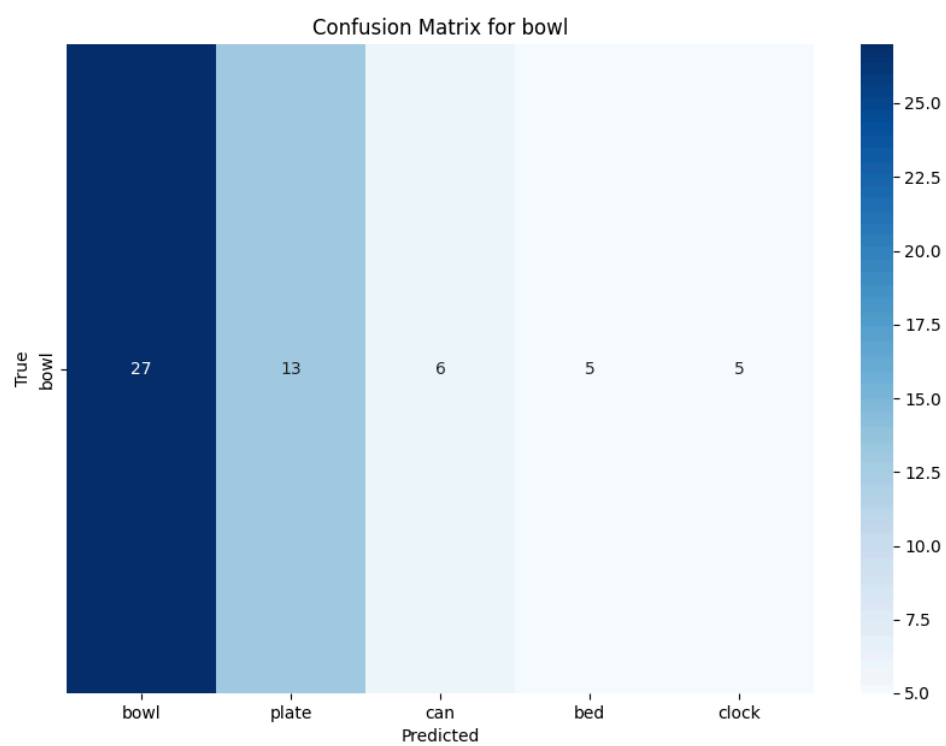
Tak samo jak wcześniej pokazane są macierze pomyłek, gdy oczekiwana była pewna klasa a model sklasyfikował obrazek do innej etykiety. Macierzy są ukazane dla 3 najlepiej przewidywanych klas oraz trzech najgorzej przewidywanych klas.

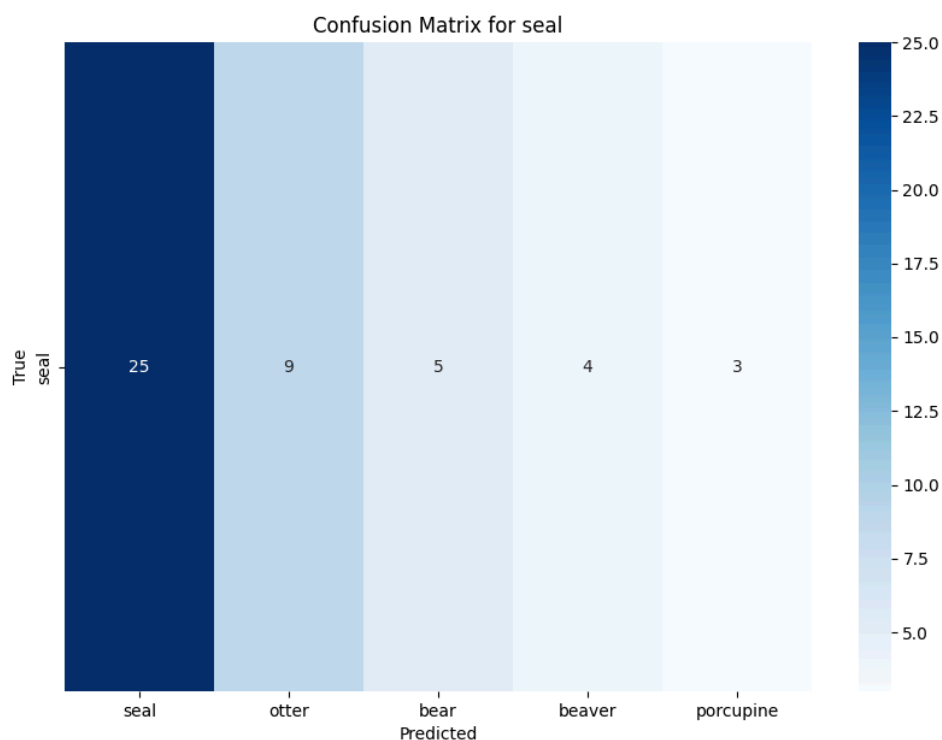
Najlepiej przewidywane klasy: **Wardrobe (92%), Orchid(91%), Road(90%)**





Najgorzej przewidywane klasy: **Bowl(27%), Otter(25%), Seal(25%)**





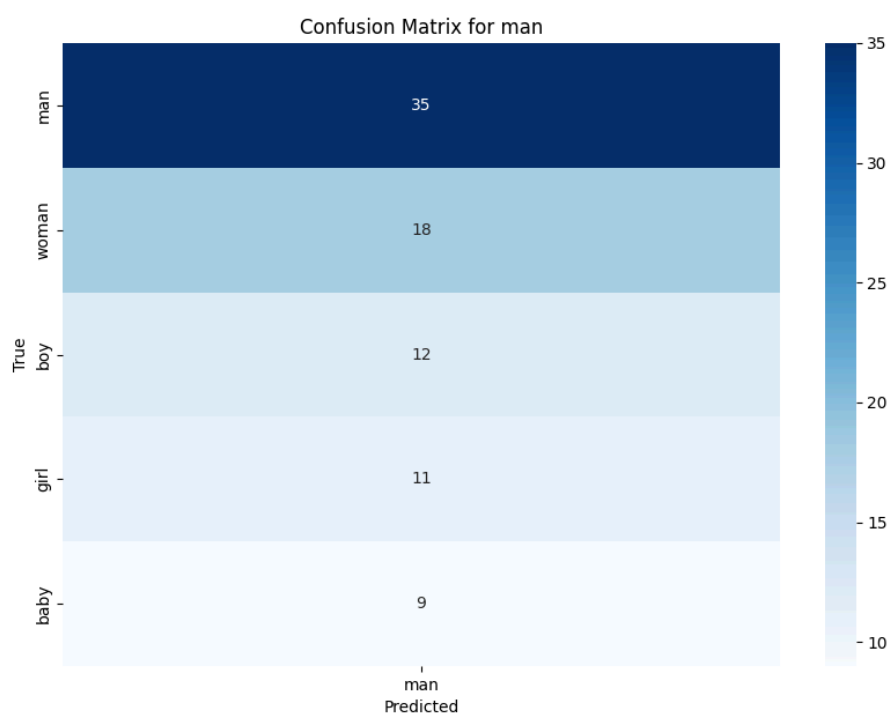
Metryki oceny

Precyzja: 0.656

Była liczona w ten sam sposób jak dla najlepszego modelu.

Klasa z najlepszą precyzją: skunk (0,90)

Klasa z najmniejszą precyzją: Man (0,27) - na wykresie poniżej można zobaczyć, że model przewiduje często fałszywie pozytywnie klasę Man:



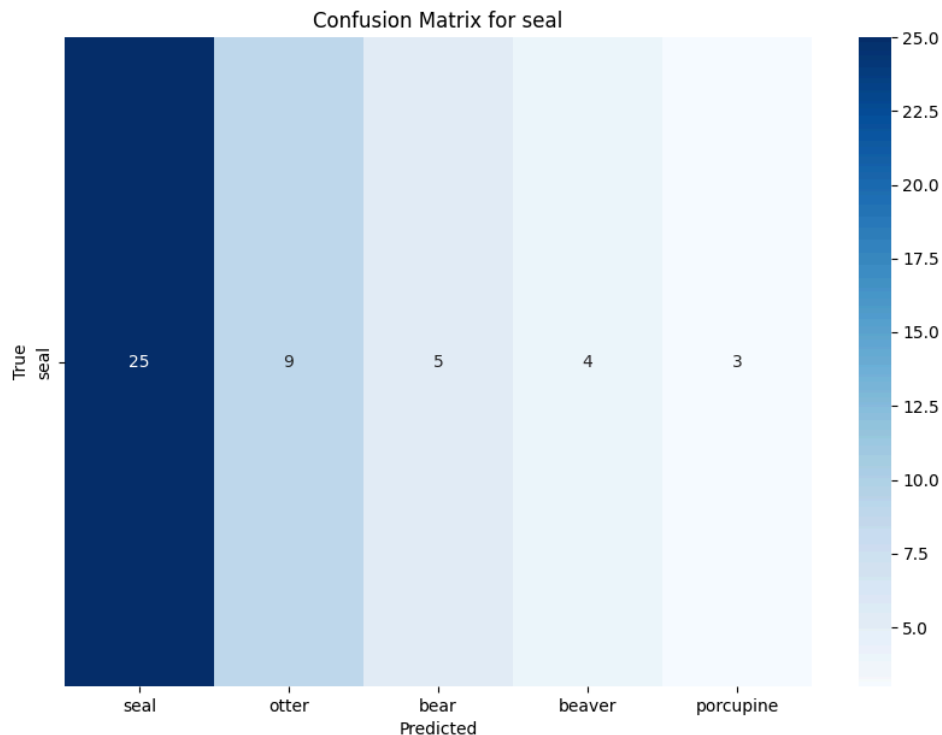
Czułość: 0.643

Czułość była obliczana tak samo jak dla najlepszego modelu.

Klasa z najlepszą czułością: Wardrobe (0,91)

Klasy z najmniejszą czułością: Seal (0,25) - jak widać na wykresie poniżej dla klasy

Seal model przewiduje dużo fałszywie negatywnych(FN) klas:

**F1-Score: 0.643**

Średnia harmoniczna precyzji i czułości dla tego modelu również wyszła podobna do wartości precyzji i czułości. Zatem model osiąga zrównoważoną wydajność i nie faworyzuje żadnej z miar, co jest też zasługą dobrze zbalansowanych danych, gdyż każda klasa w zbiorze CIFAR100 ma taką samą liczbę obrazków.