Projekt – sieci neuronowe	Data złożenia projektu: 27.05.2022
Numer grupy projektowej: CWP 4	Imię i nazwisko I: Szymon Musiał Imię i nazwisko II: Piotr Wilkosz

TSR – System rozpoznawania znaków (polskich)

1. Opis problemu i danych

W ramach projektu poruszana jest dziedzina klasyfikacji obrazów. Klasyfikowanymi obrazami są polskie znaki drogowe. Celem jest nauczenie sieci poprawnej klasyfikacji zdjęcia znaku do symbolu znaku.

Zdjęcia z kamery przedniej samochodu, skalsyfikowane przez model mogą być pomocą dla kierowcy.

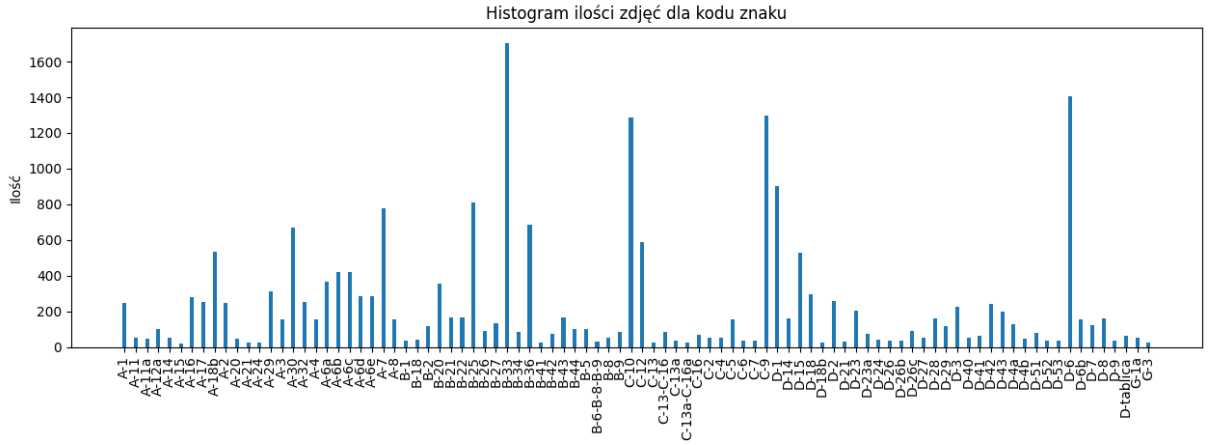
Analizowany zbiór danych składa się z 21032 rekordów. Znajdziemy w nim 92 zmienne wejściowe. Liczność wybranych zmiennych zaprezentowana jest na poniższym histogramie. Problematyczne staje się przedstawienie podstawowych statystyk ze względu, iż rozpatrywany jest problem klasyfikacji obrazów.

Pod względem liczności rekordów w klasie możemy zauważyć, że maksymalna ilość obrazów w klasie wynosi 1706 dla klasy 38 o symbolu B-33 – Ograniczenie prędkości.

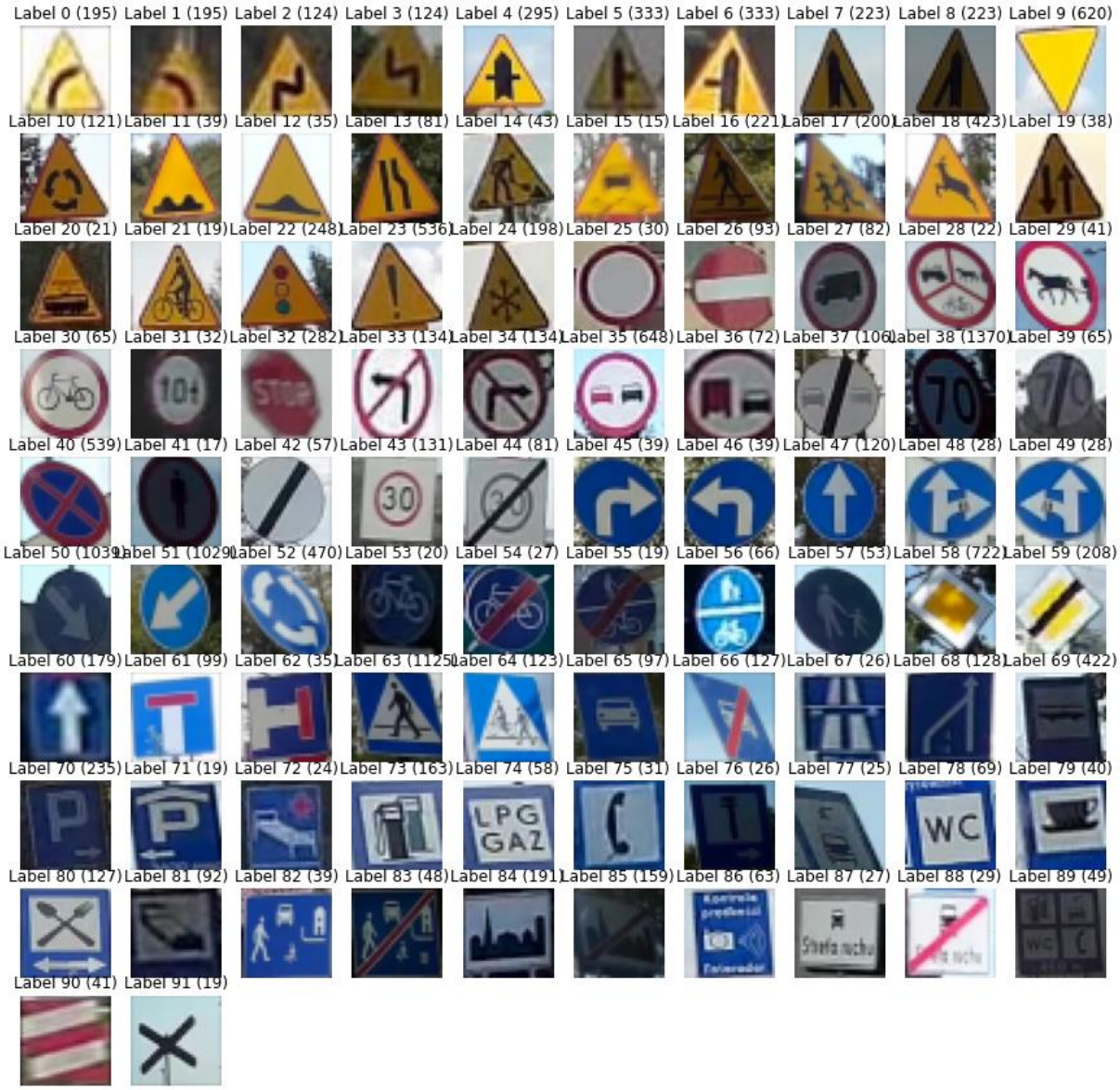
Przechodząc do analizy klasy o minimalnej ilości rekordów możemy wyróżnić klasę 15 o symbolu A-– Śliska jezdnia o liczności równej 20.

Średnia ilość rekordów na klasę wynosi 228.61 rekordy.

Odchylenie standardowe wynosi 319.02.



Wykres 1. Ilość zdjęć dla danego kodu znaku



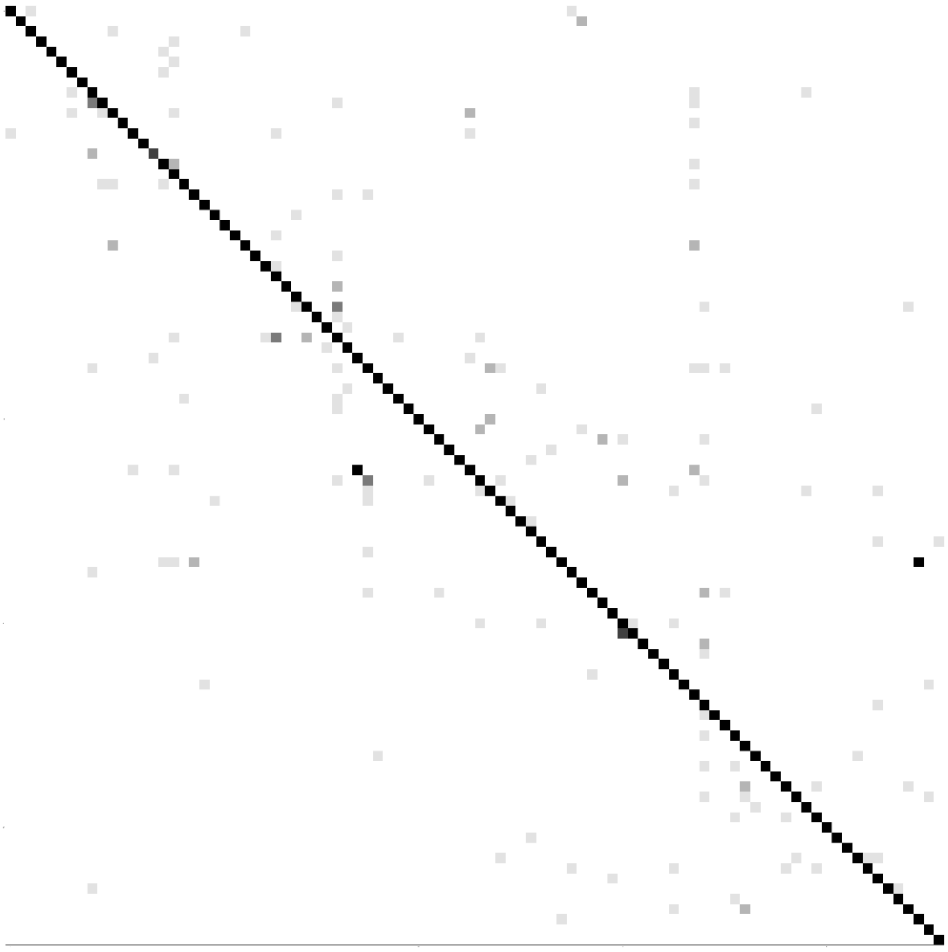
Rys 1. Przedstawienie klas

2. Obróbka danych

Otrzymany zbiór posiadał już podział na testujący oraz treningowy. Ilość rekordów w poszczególnych zbiorach rozkładała się w proporcji 80:20. Ze zbioru treningowego 10% zdjęć zostało przeniesione do zbioru walidującego używanego przy trenowaniu.

Ponadto zbiór testujący po wytrenowaniu sieci został użyty w celi walidacji dokładności.

Poniżej zaprezentowano macierz po odfiltrowaniu zdjęć złej jakości, na której to osi X znajduje się klasa przypisana do zdjęcia testującego a do osi Y klasyfikacja modelu. Odchyłki od głównej przekątnej stanowią błędy dopasowania. Kolor elementów macierzy jest szary dla 5 elementów



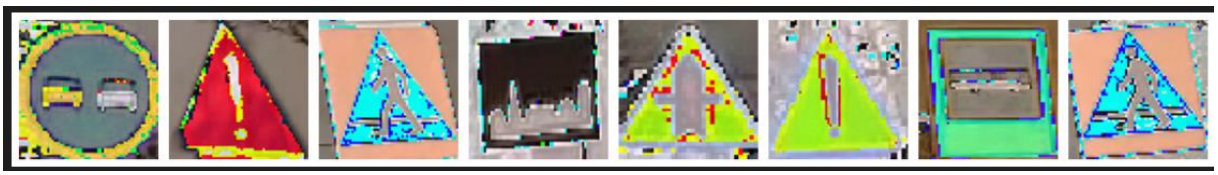
Rys 2. Graficzna macierz dopasowania modelu

Macierz ta była decyzyjną w procesie filtrowania, a zaprezentowana powyżej jest już wynikową.

Ponadto zastosowany ImageDataGenerator od razu dokonuje powiększenia zbioru stosując np. przesunięcie, skalowanie, rotacje i tym podobne przekształcenia.

W celu modyfikacji zdjęć podjęto próby wykorzystać to VGG16 oraz VGG19, przekazany jako argument ImageDataGenerator reprocessing_function=tf.keras.applications.vgg16.preprocess_inpu

Jednakże spowodowało to spadek dokładności modelu, dlatego zrezygnowano z niego. Gdyby użyto przetrenowanego wcześniej modelu skorzystanie z VGG lub podobnych modelów byłoby wręcz koniecznością ze względu na dostępność danych w tych modelach



Rys 3. Zastosowany filtr VGG16

Dodatkowo sprawdzono filtr szum Gaussa, który również nie przyczynił się do poprawy wyników.

Rys 4. Dane – obrazy wejściowe

3. Opis zastosowanych sieci neuronowych

W projekcie zastosowano konwolucyjną sieć neuronową

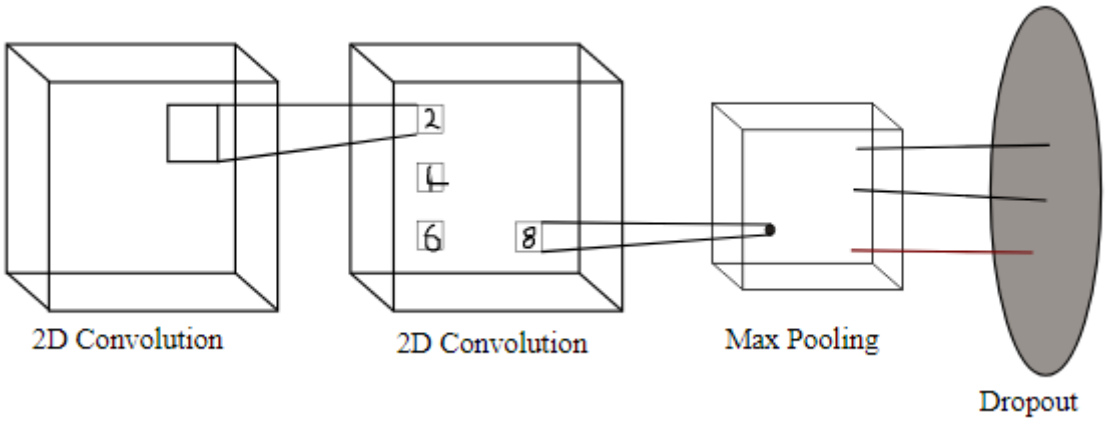
Ilość epok uczenia została empirycznie dobrana na 25, mniejsza liczba powodowała niską dokładność a większa powodowała przetrenowanie co skutkowało rozbieżnością na poziomie kilkudziesięciu procent w testach dokładności na danych testujących i walidujących podczas uczenia modelu

Do uczenia użyto biblioteki Keras który jako backend wykorzystuje bibliotekę Tensorflow w wersji 2.5, który natomiast wykorzystuje kartę graficzną w celu przyspieszenia uczenia. Oprócz tego w celu wizualizacji uzyskanych wyników używamy biblioteki matplotlib. Do konstrukcji macierzy konfuzji używamy możliwości oferowanych przez bibliotekę sklearn. Do operacji matematycznych wykorzystujemy bibliotekę numpy.

Schemat sieci podzielona na dwie części: część ekstrakcji cech oraz klasyfikacji. Część wyodrębniająca cechy jest zapętlona dwa razy zanim trafi do części klasyfikacji.

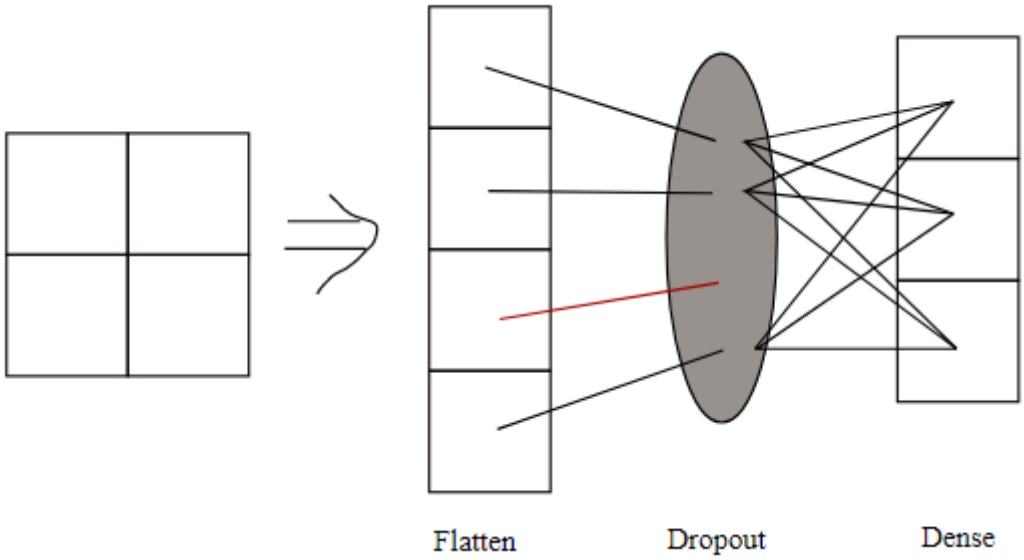
Dwie pierwsze połączone warstwy konwolucji mają 32 filtry oraz rozmiar jądra 5x5. Natomiast druga iteracja posiada tych filtrów 64 a jądro ma rozmiar 3x3. Warstwy każdorazowo mają funkcję aktywacji ReLU (GeLU przyniosło gorsze wyniki).

Warstwa max pooling ma rozmiar 2x2, a współczynnik losowo usuwanych połączeń wynosi 25% zapobiegając przetrenowaniu



Rys 4. Wyodrębnianie cech

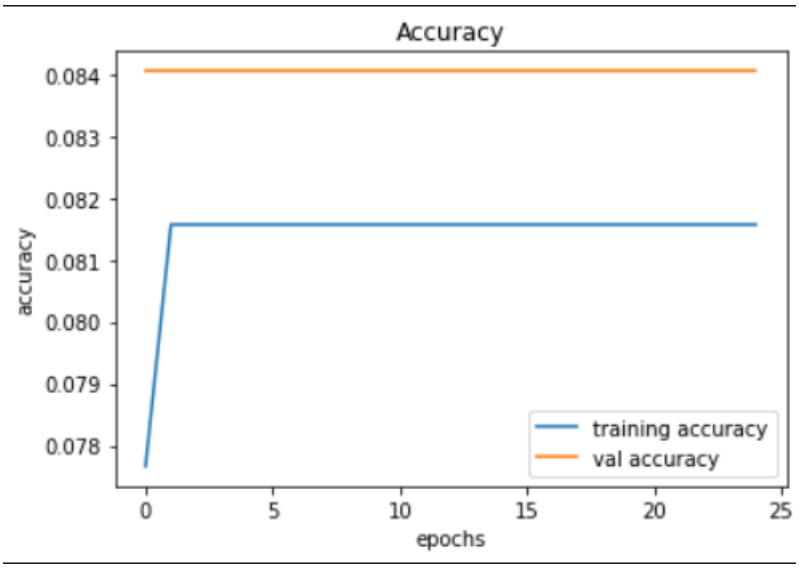
Następnie dokonywana jest klasyfikacja



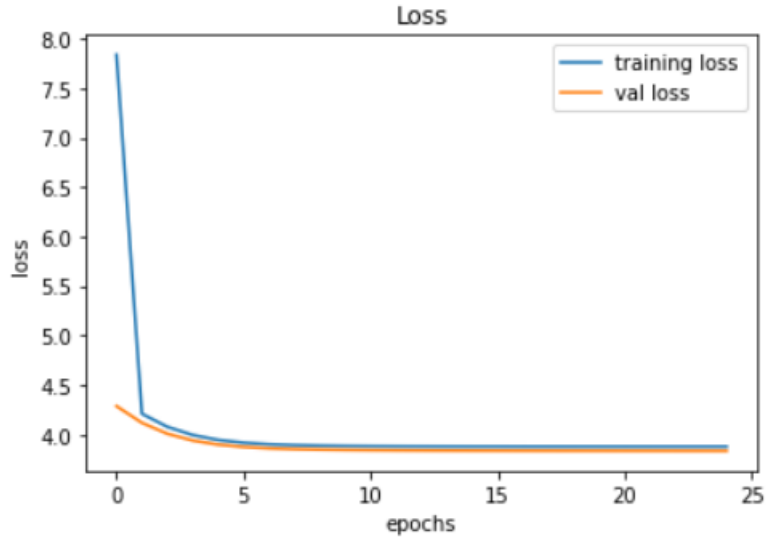
Rys 5. Klasyfikacja

Wyjściowa warstwa w poprzednim procesie jest poddawana transformacji na 1 wymiar (Flatten). Następnie taki sam procent połączeń, jak poprzednio, jest losowo usuwana (Dropout) i ostatecznie łączona każdy z każdym w warstwie Dense

Wyniki dokładności modelu składającej się z pojedynczej warstwy konwolucji

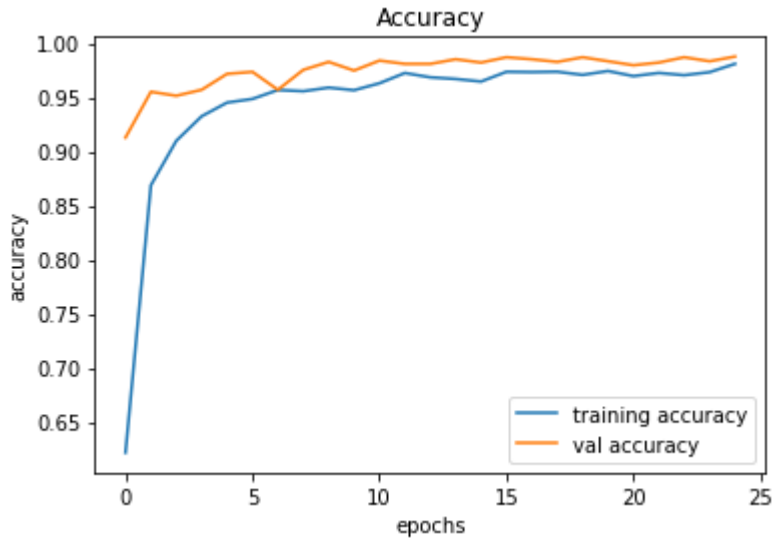


Rys 6. Dokładność modelu składającego się z pojedynczej warstwy

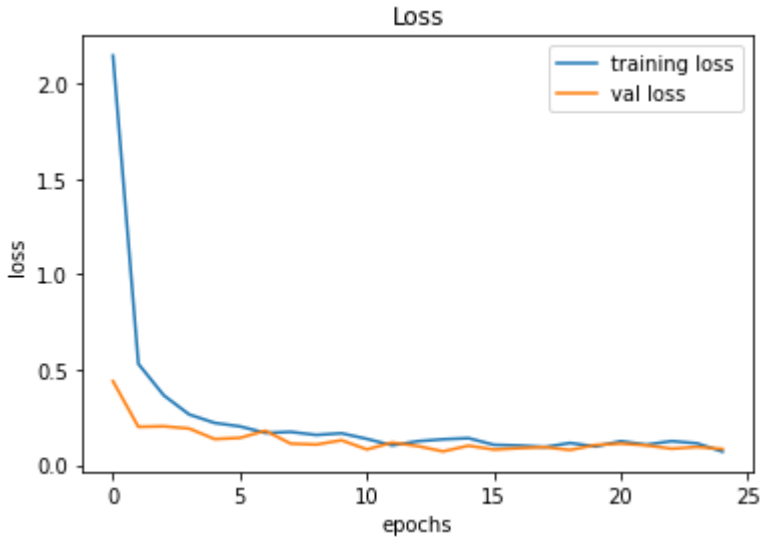


Rys 7. Kara w modelu składającym się z jednej warstwy

Z tego też powodu dokonano modyfikacji sieci na taką jaka jest przedstawiona na Rys. 4 oraz Rys. 5 uzyskując poniższe dokładności



Rys 8. Dokładność modelu składającego się z wielu warstw



Rys 9. Kara w modelu składającym się z wielu warstw

4. Dyskusja wyników oraz wnioski

Uważamy, że uzyskane wyniki są satysfakcjonujące. Specyfika rozpatrywanego przez nas problemu jest problemem skomplikowanym. Jakość używanych zdjęć różnie nie jest rewelacyjna. Liczność poszczególnych klas, ilość znaków drogowych w kategoriach bardzo się różni. Czasem zdjęć jednego znaku jest 20 a czasem 1200. Wynika to z czystej specyfiki rozpatrywanego problemu. Naturalnym zjawiskiem pod względem kraju, w którym się znajdujemy jest, że na drodze częściej znajdziemy znaki ograniczające maksymalną dozwoloną prędkość niż znaki na przykład ostrzegające o przeprawie promowej lub o zakazie wjazdów motorowerów. Z tego też powodu napotykamy się problemem wysokiego odchylenia standardowego wynoszącego 319. W związku z tym uważamy, że uzyskany przez nas współczynnik dokładności 95.6 jest satysfakcjonujący biorąc pod uwagę specyfikę problemu.

Dalszym etapem rozwoju projekty mogłoby być zastosowanie mechanizmu detekcji znaków drogowych w czasie rzeczywistym dzięki czemu możliwy byłoby zbudowanie systemu informującego kierowcę o drodze.