

Projekt zaliczeniowy

Seminarium bioinformatyczne

Daria Plewa

Michał Humiński

Dominik Lisiecki

Bioinformatyka

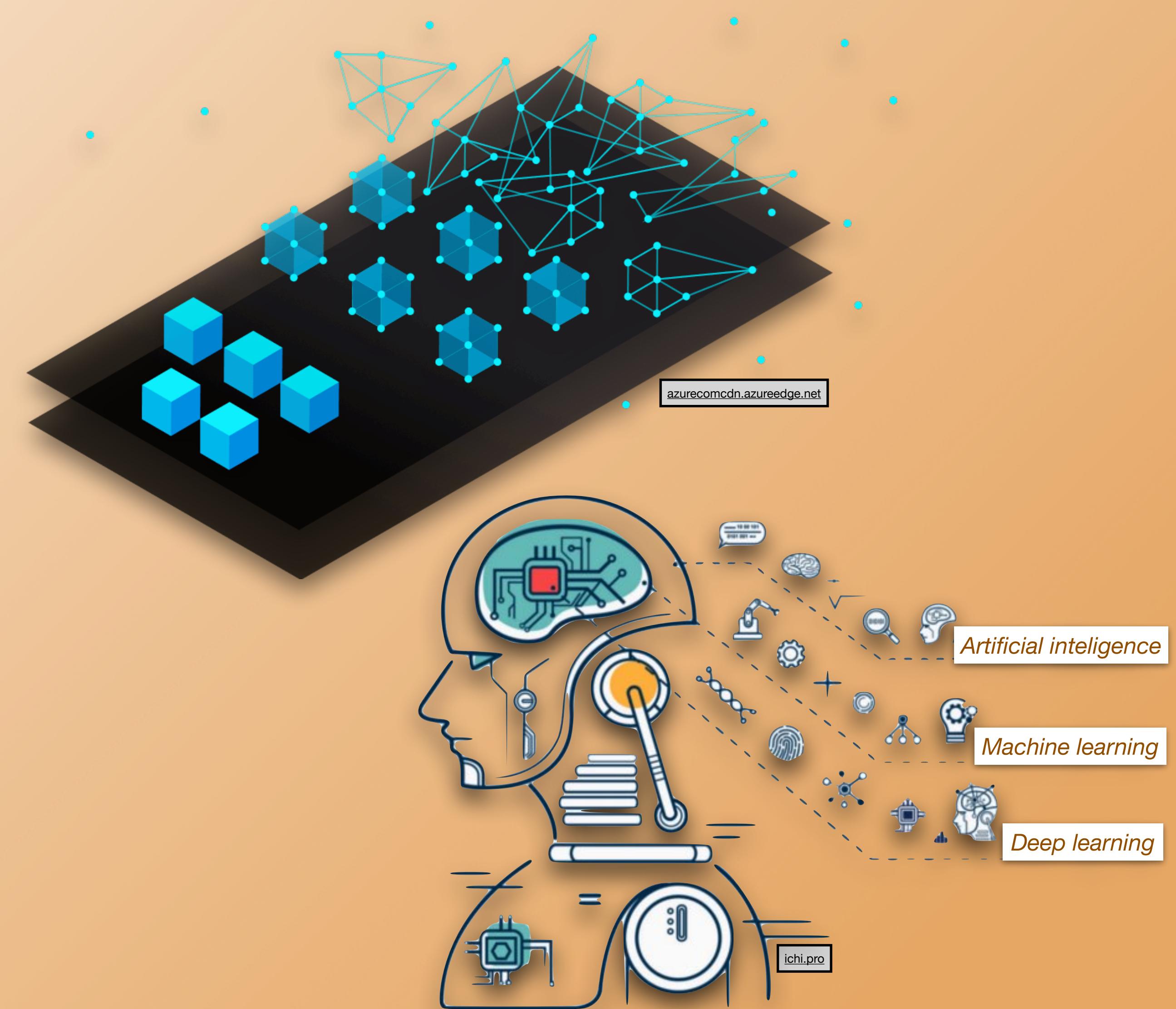
studia licencjackie rok 3 2021/2022



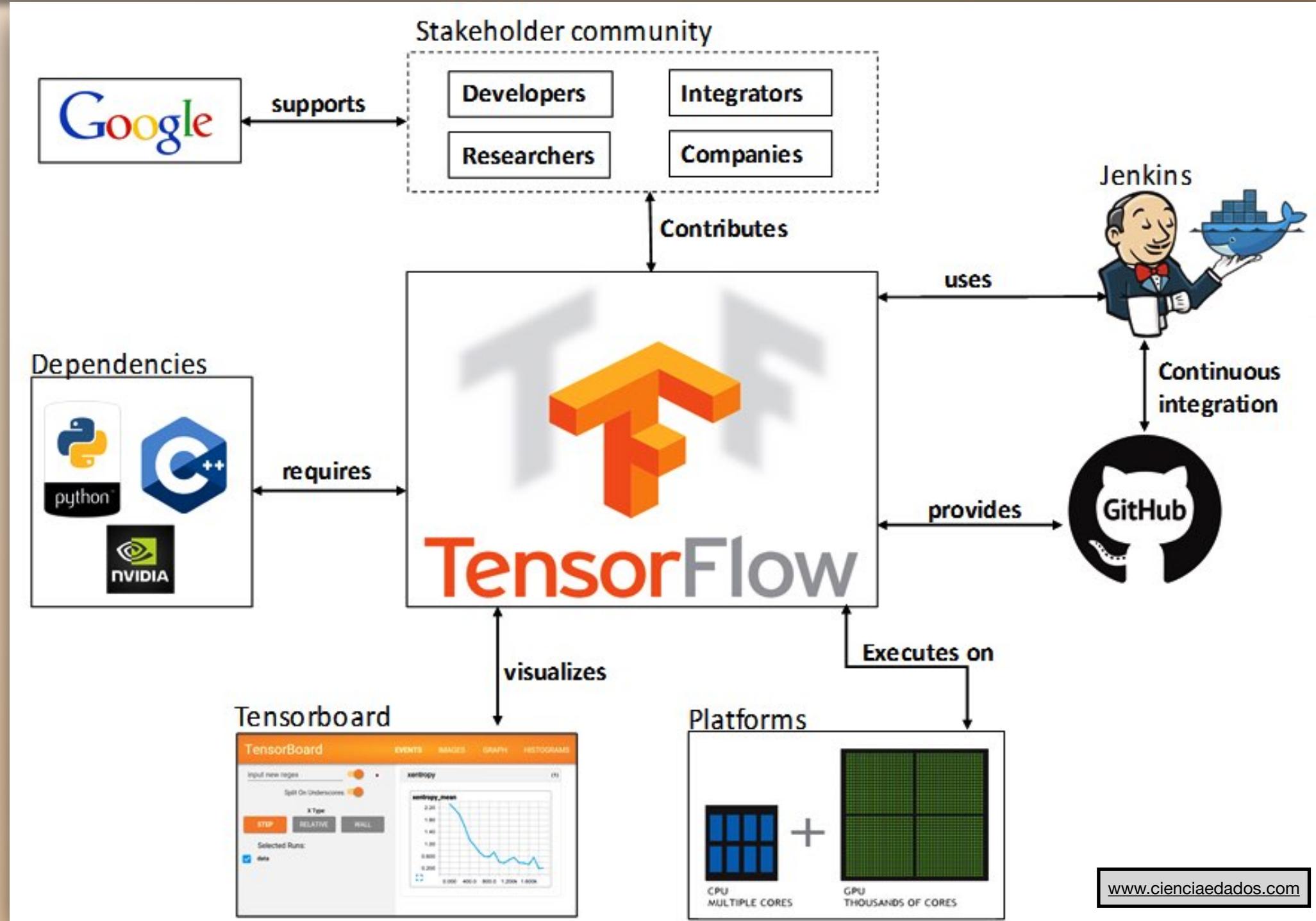
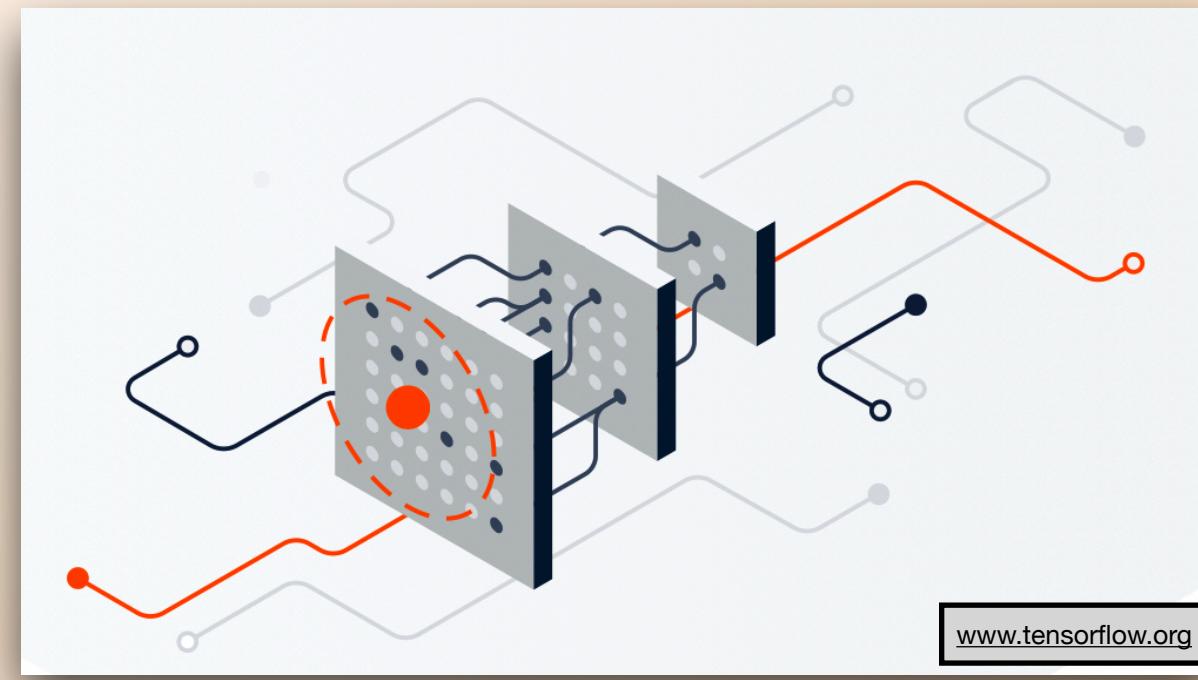
Część teoretyczna

Uczenie maszynowe

- ▶ Jest to używanie matematycznych modeli danych w celu **ułatwienia komputerowi uczenia się** bez bezpośrednich instrukcji.
- ▶ **Dokładność wyników uczenia maszynowego** zwiększa się wraz z upływem czasu i wzrostem ilości danych.



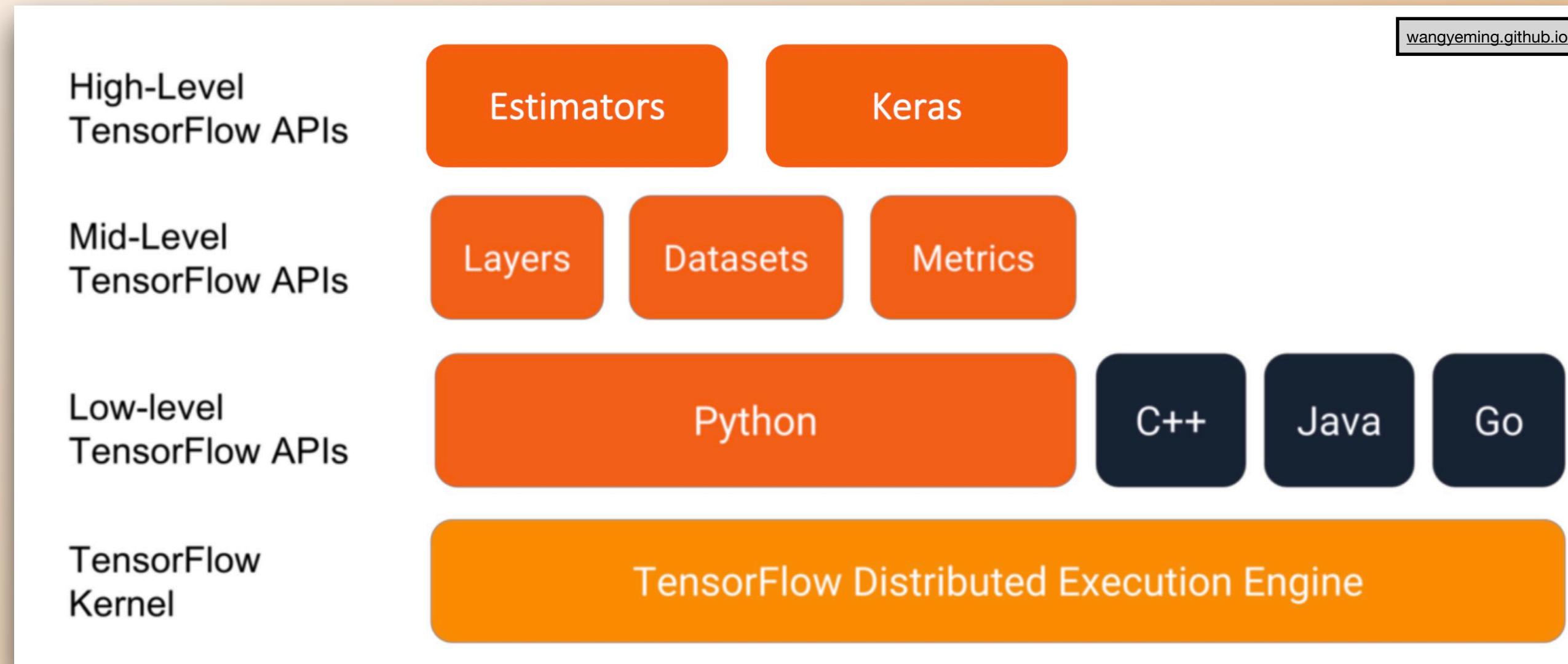
Biblioteka TensorFlow



pl.wikipedia.org

- ▶ **TensorFlow** to bardzo rozbudowana otwartoźródłowa biblioteka programistyczna do *machine learningu*, która cechuje się wysoką wydajnością i skalowalnością.
- ▶ **Elastyczna architektura** umożliwia wdrożenie obliczeń na jednym lub większej liczbie procesorów (CPU) lub kart graficznych (GPU) bez konieczności ponownego pisania kodu.

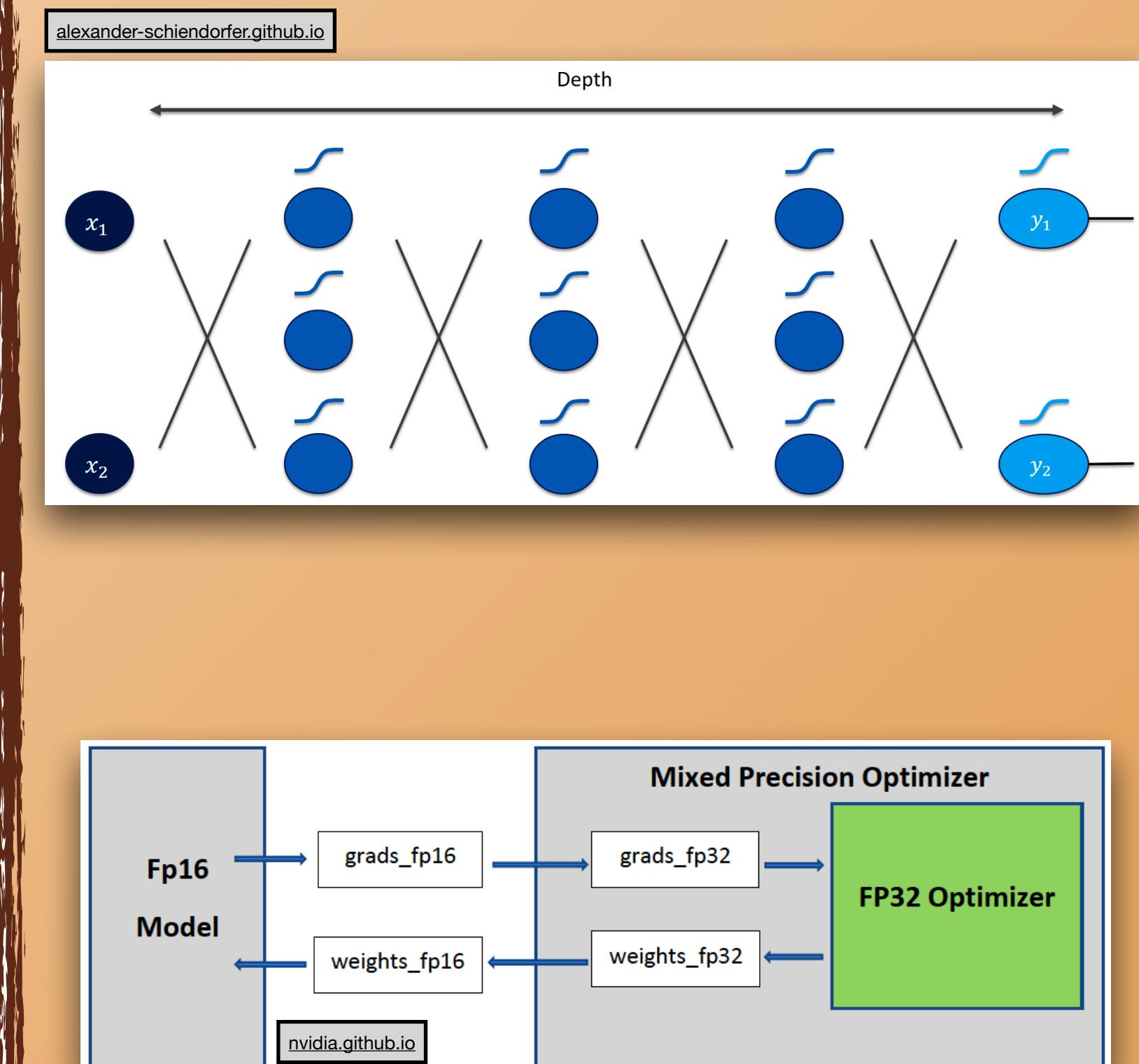
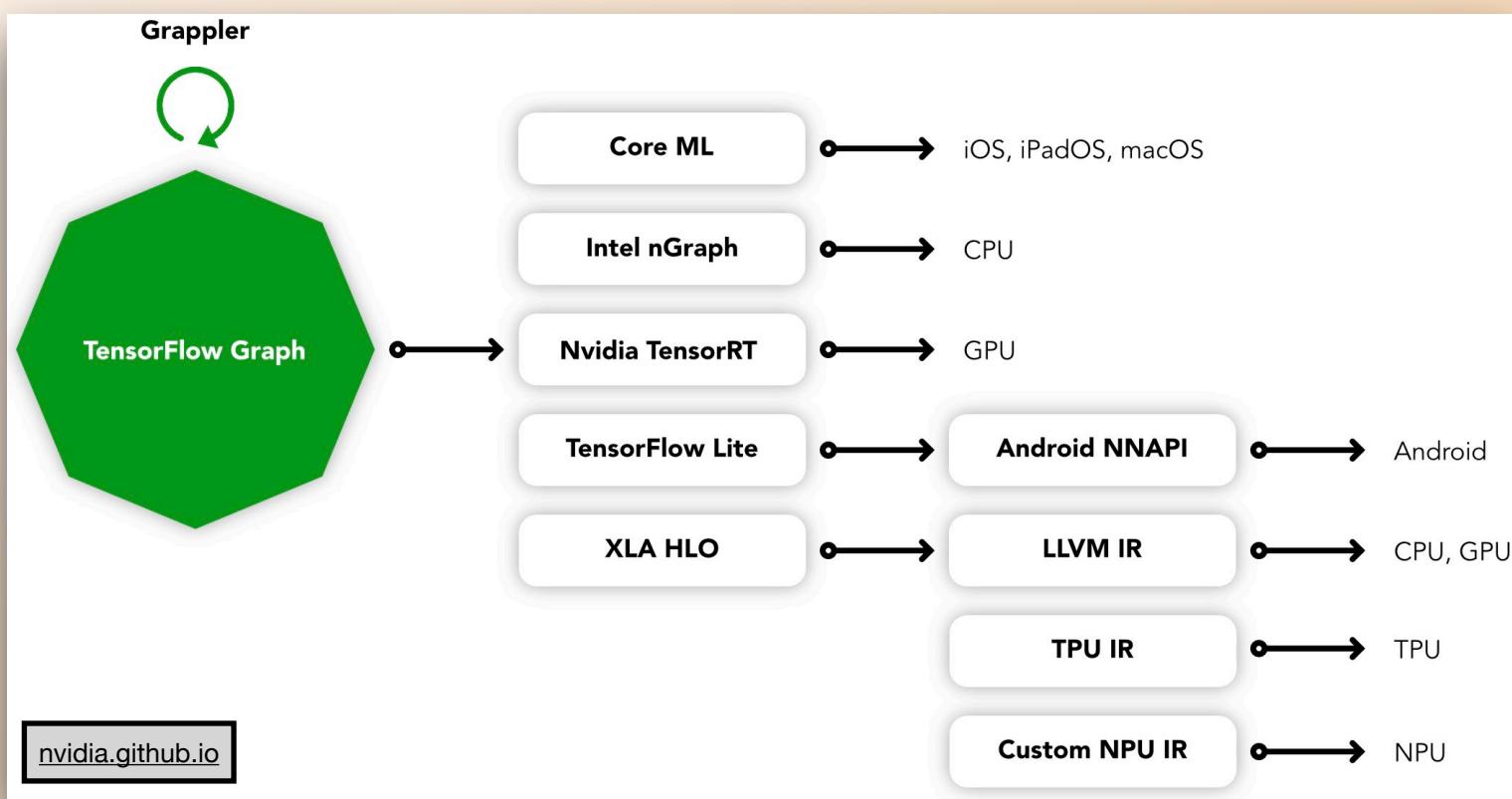
Biblioteka TensorFlow



- ▶ **Biblioteka składa się z kilku modułów:**
 - rozproszony silnik wykonawczy (*distributed execution engine*),
 - frontendy,
 - warstwa API,
 - wysokopoziomowe API,
 - biblioteki z gotowymi modelami do użycia.

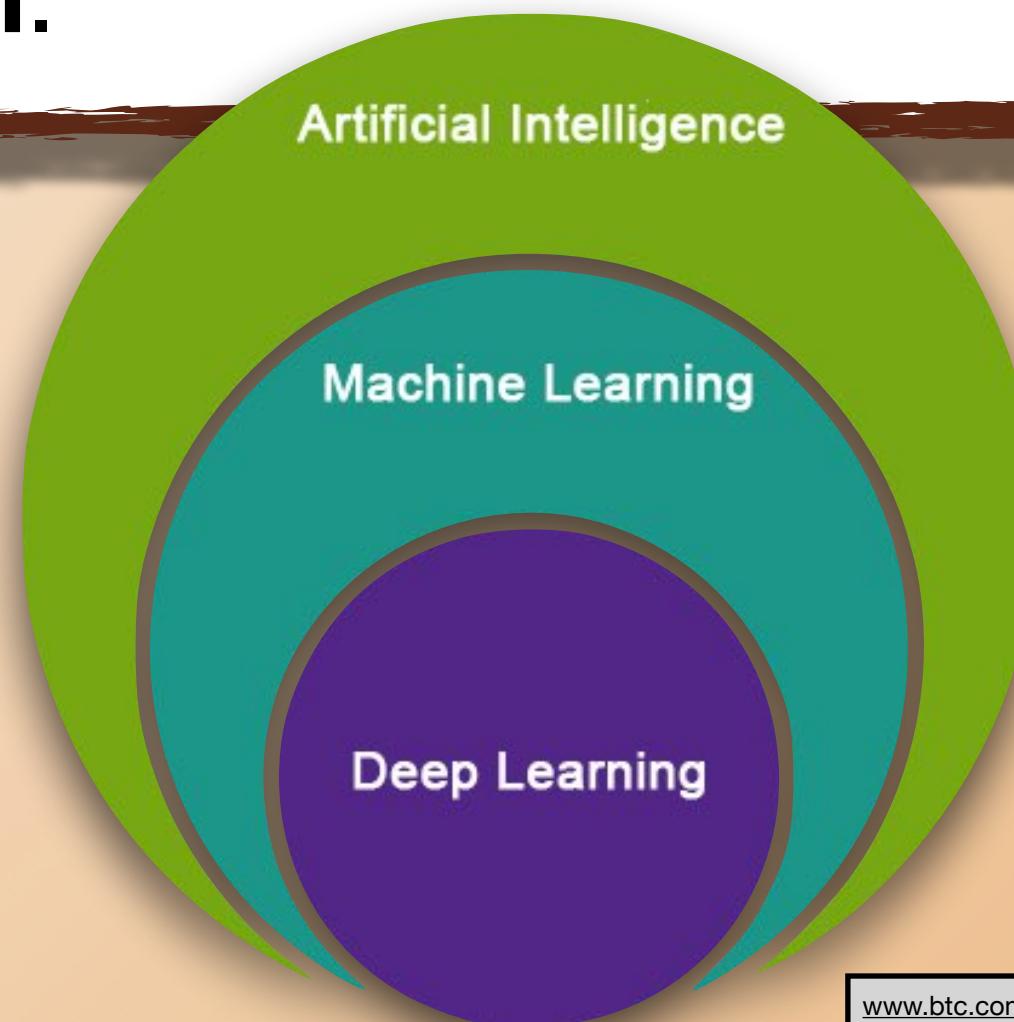
TensorFlow - cechy

- ▶ **Bibliotekę TensorFlow charakteryzuje:**
 - ❖ automatyczne różnicowanie,
 - ❖ *eager execution*,
 - ❖ dostępność,
 - ❖ wykorzystanie funkcji strat,
 - ❖ metryki oceny,
 - ❖ *TF.nn*,
 - ❖ optymalizatory.



TensorFlow - zastosowanie

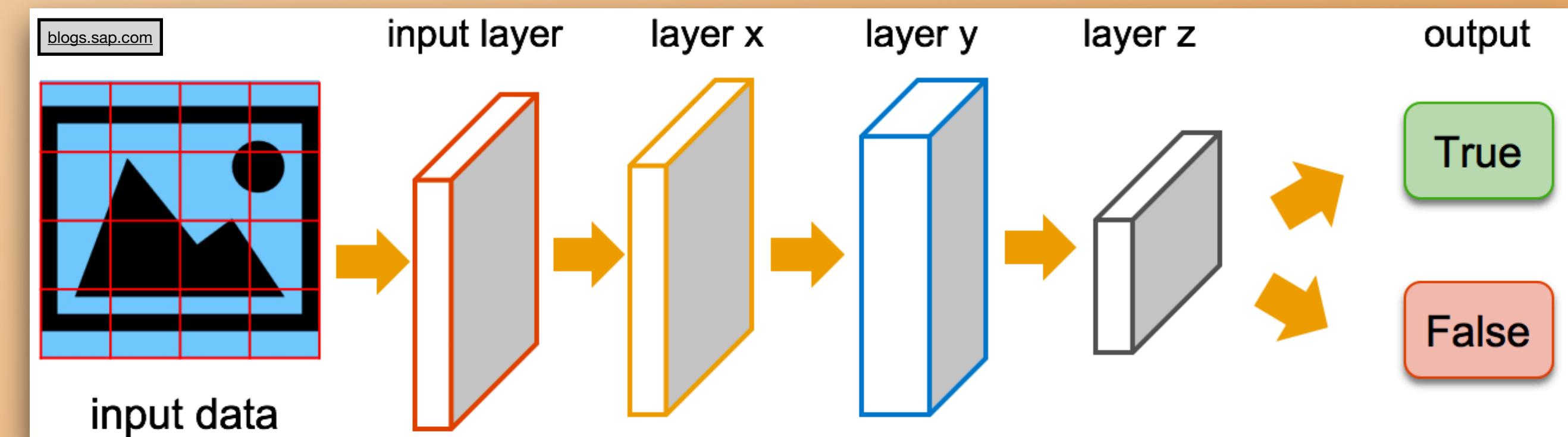
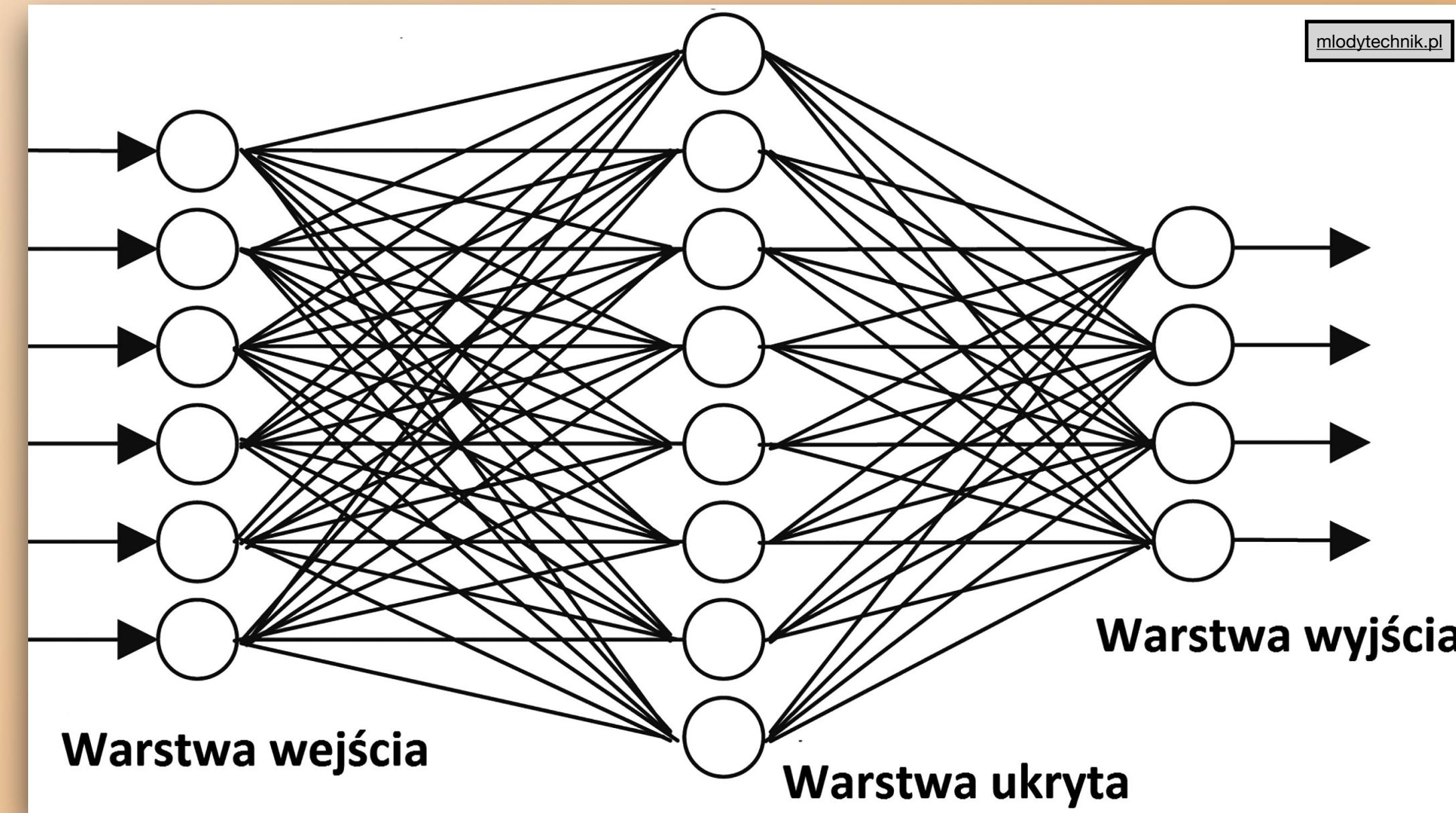
- ▶ TensorFlow został zaprojektowany, aby realizować projekty badawcze w dziedzinie **maszynowego uczenia się** oraz **głębokich sieci neuronowych**.
- ▶ Umożliwia optymalne uczenie się modeli wymagających **dużej ilości danych**.



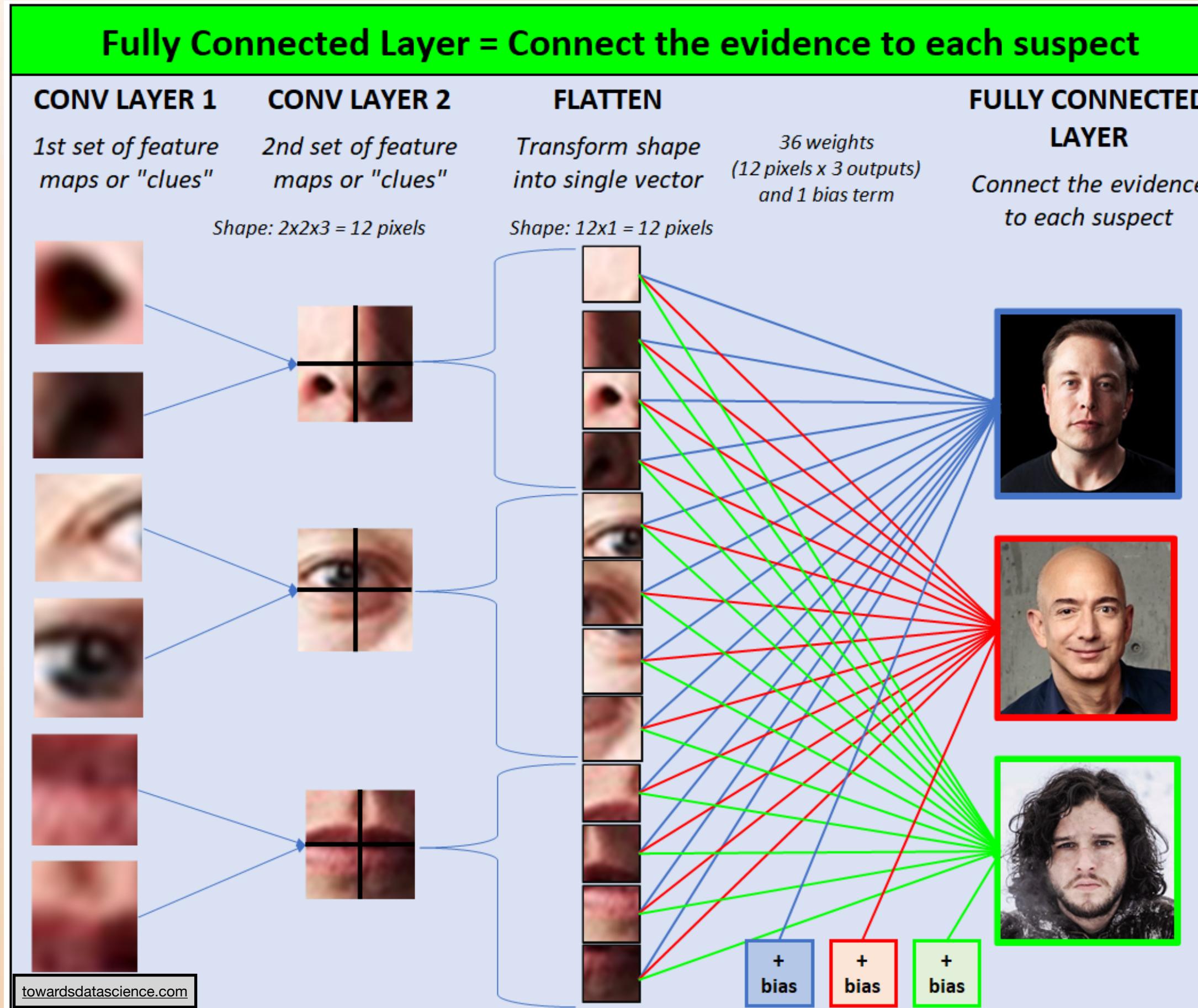
- ▶ TensorFlow służy również do:
 - rozpoznawania mowy,
 - wyszukiwania głosowego,
 - detekcji błędów,
 - wykrywania języków,
 - określania niebezpieczeństw w mediach społecznościowych,
 - rozpoznawania obrazów,
 - detekcji video.

CNN - splotowe sieci neuronowe

- ▶ **Splotowa (koewolucyjna) sieć neuronowa** to pewien rodzaj układu sztucznych neuronów lub symulatorów neuronów, który ma działać w określony sposób.
- ▶ To **sztuczna sieć neuronowa** stworzona w celu symulacji pewnego rodzaju aktywności mózgu.



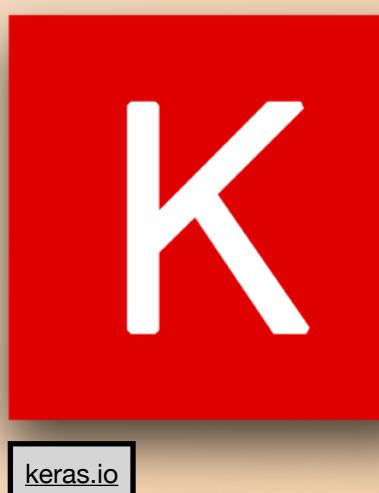
CNN - splotowe sieci neuronowe



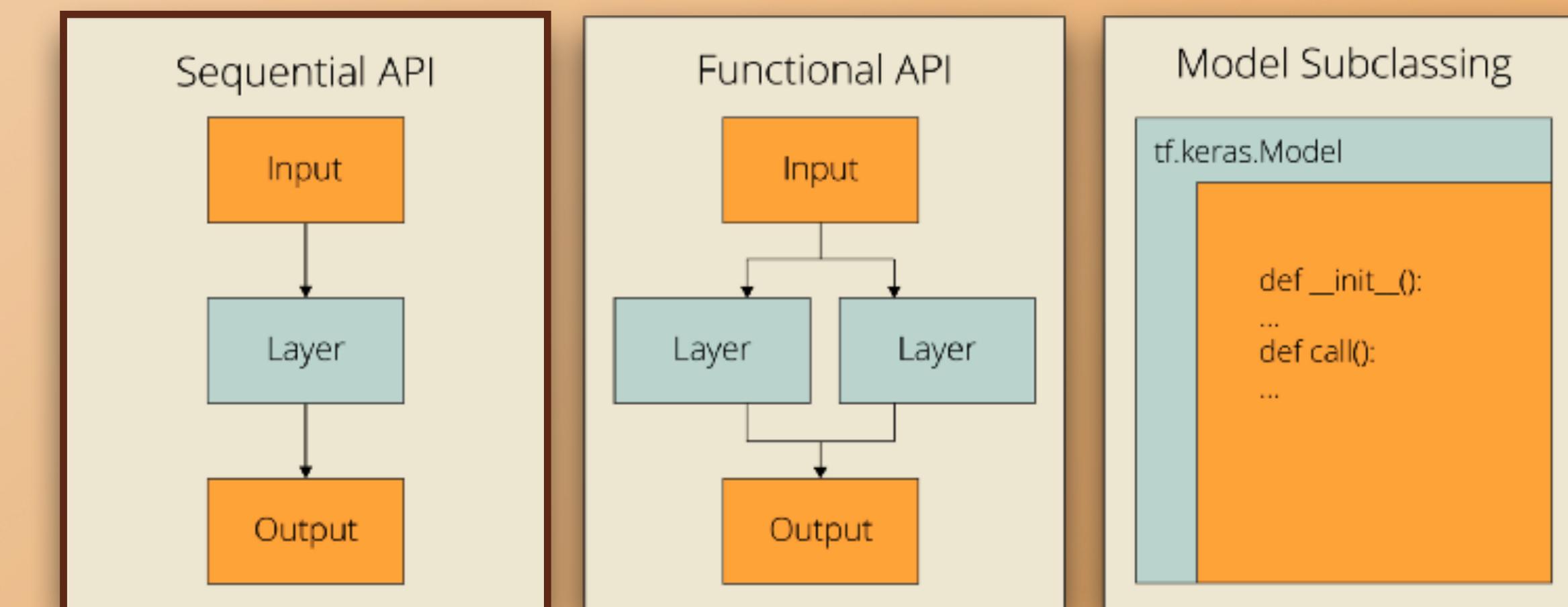
- ▶ **Technologie** wykonują niektóre **zadania**, które kiedyś były zarezerwowane **wyłączne** dla widzenia biologicznego, przykładowo rozpoznawanie twarzy.
- ▶ Jednym z najbardziej rozpowszechnionych zastosowań splotowych sieci neuronowych jest **symulacja widzenia ludzi lub zwierząt**.

TensorFlow - CNN

- ▶ **TensorFlow** poprzez integrację z **biblioteką Keras** posiada możliwość używania modeli sekwencyjnych.
- ▶ Pozwala to nam na zmniejszenie produkowanej ilości kodu jednocześnie przy dostępie do łatwej modyfikacji warstw.



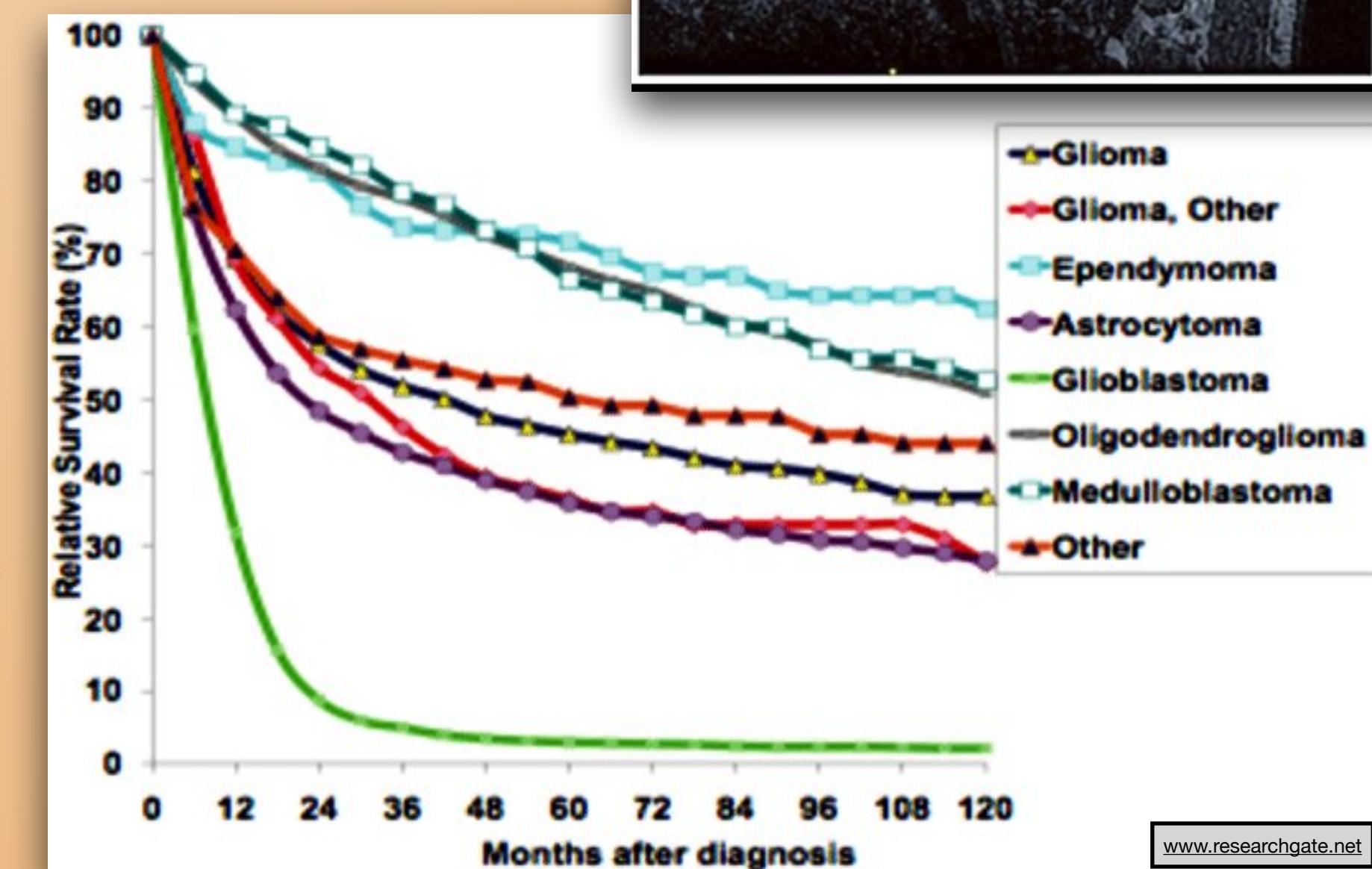
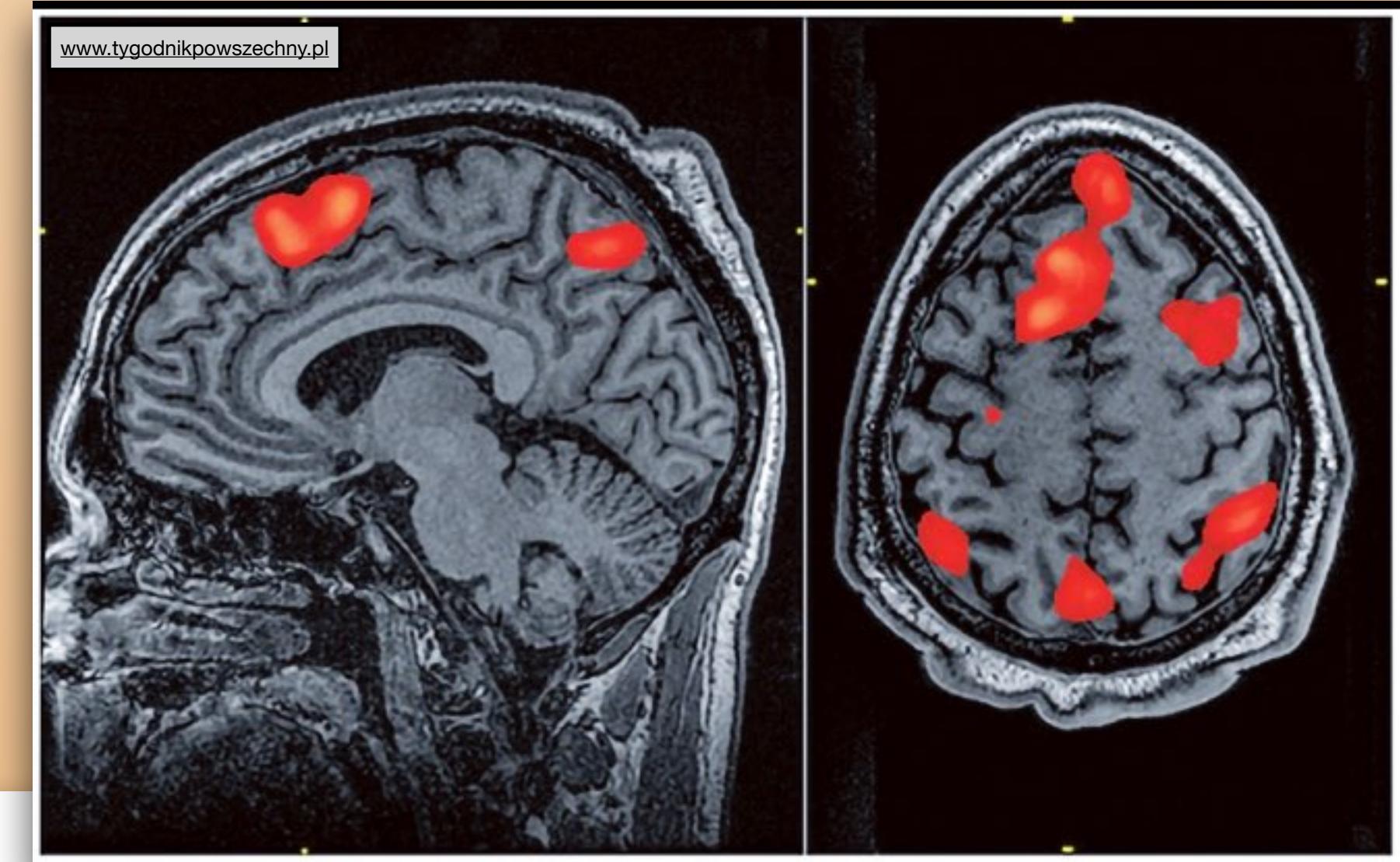
Keras



Część praktyczna

Dane

- ▶ W naszym zadaniu przetwarzamy **3200 zdjęć MRI**.
- ▶ Dotyczą one **guzów mózgu**, które są uważane za jedną z najbardziej agresywnych chorób wśród dzieci i dorosłych.
- ▶ **Guzy mózgu mają różną klasyfikację**.

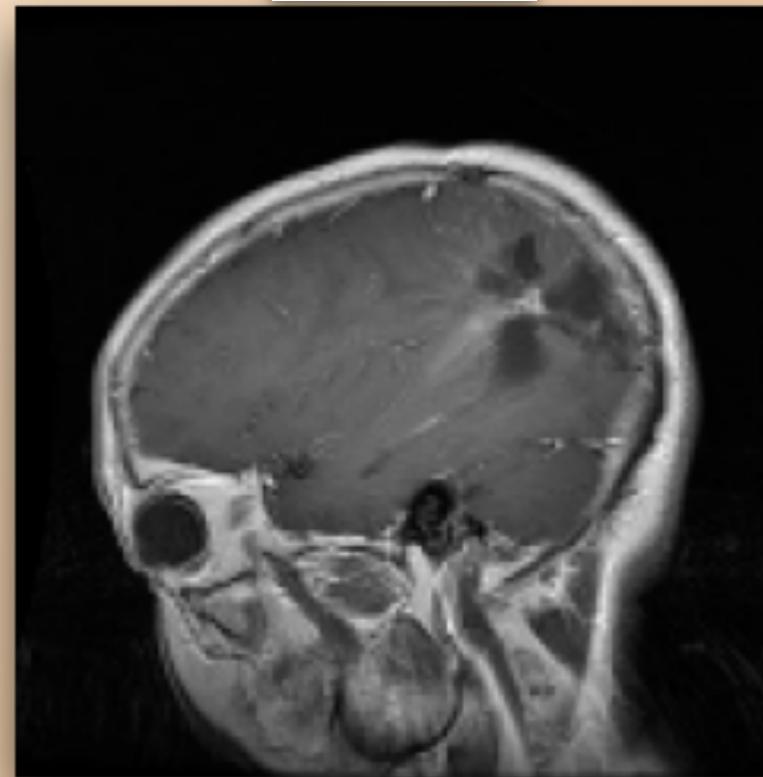


Dane

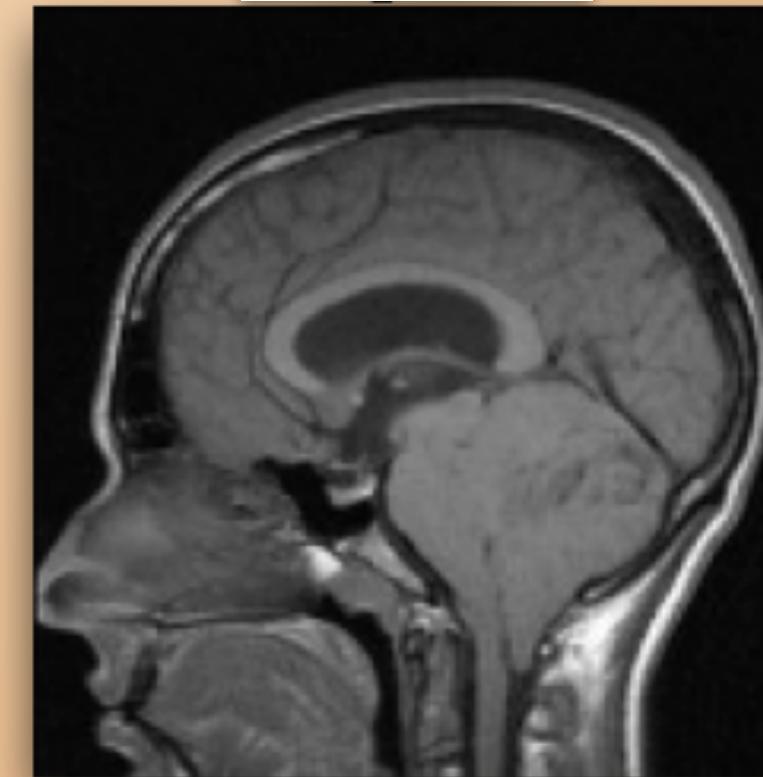
- ▶ Najlepszą techniką wykrywania guzów mózgu jest **obrazowanie metodą rezonansu magnetycznego (MRI)**.
- ▶ **Rezonans magnetyczny** w stosunku do badań radiologicznych z wykorzystaniem promieni rentgenowskich jest **nieinwazyjny**.

Sample Image From Each Label

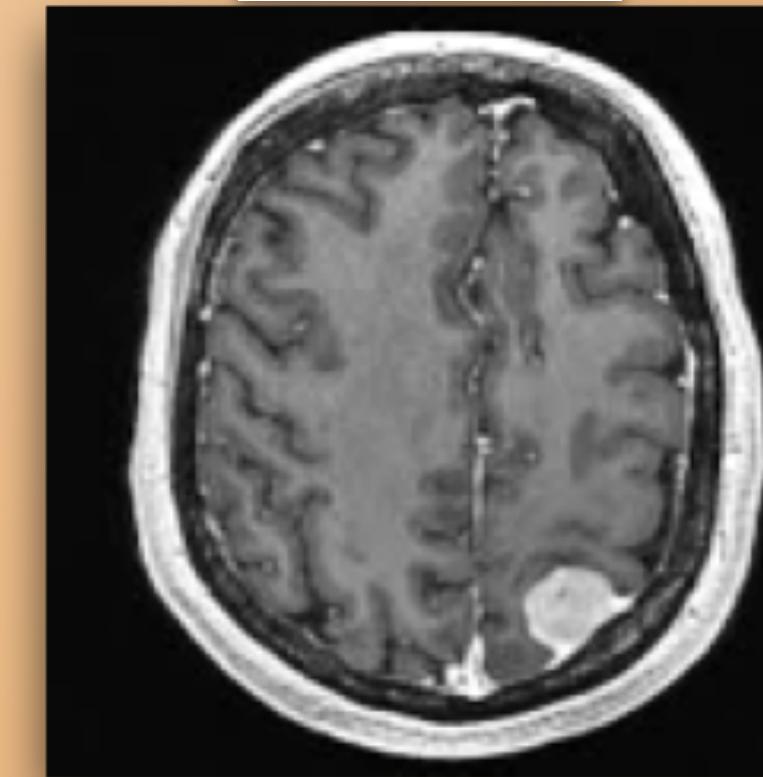
Glejak



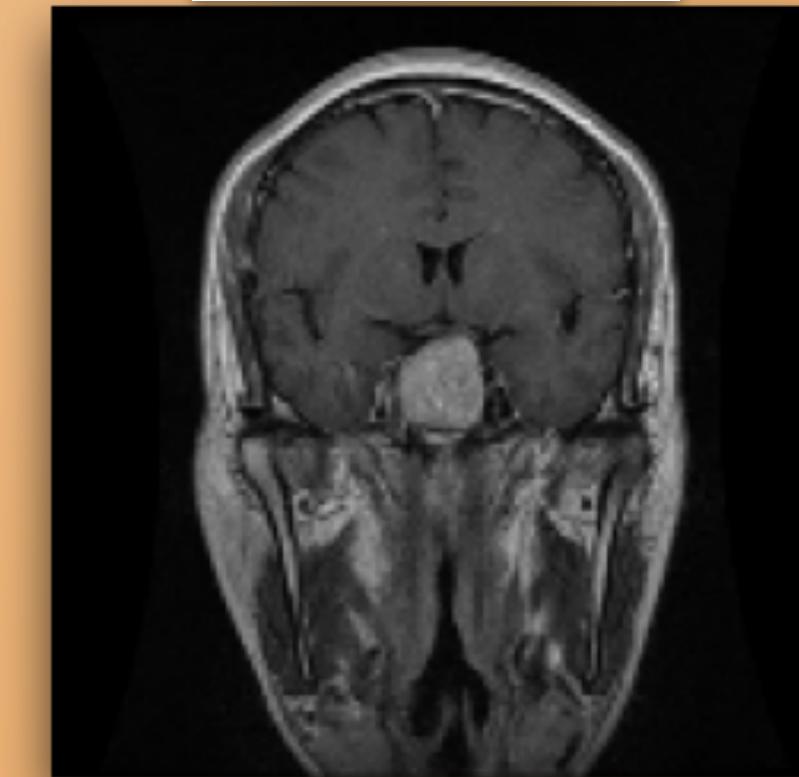
Osoba zdrowa



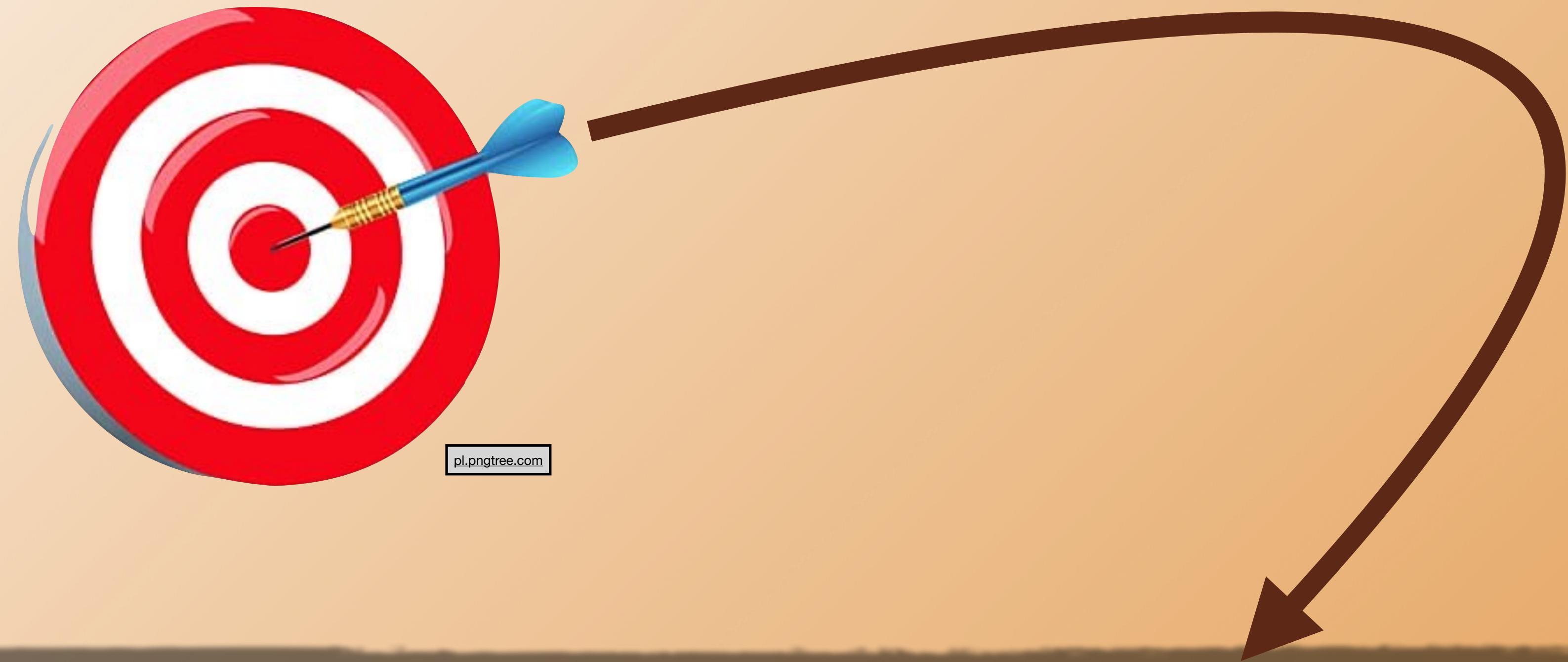
Oponiak



Nowotwór przysadki



Nasze zadanie



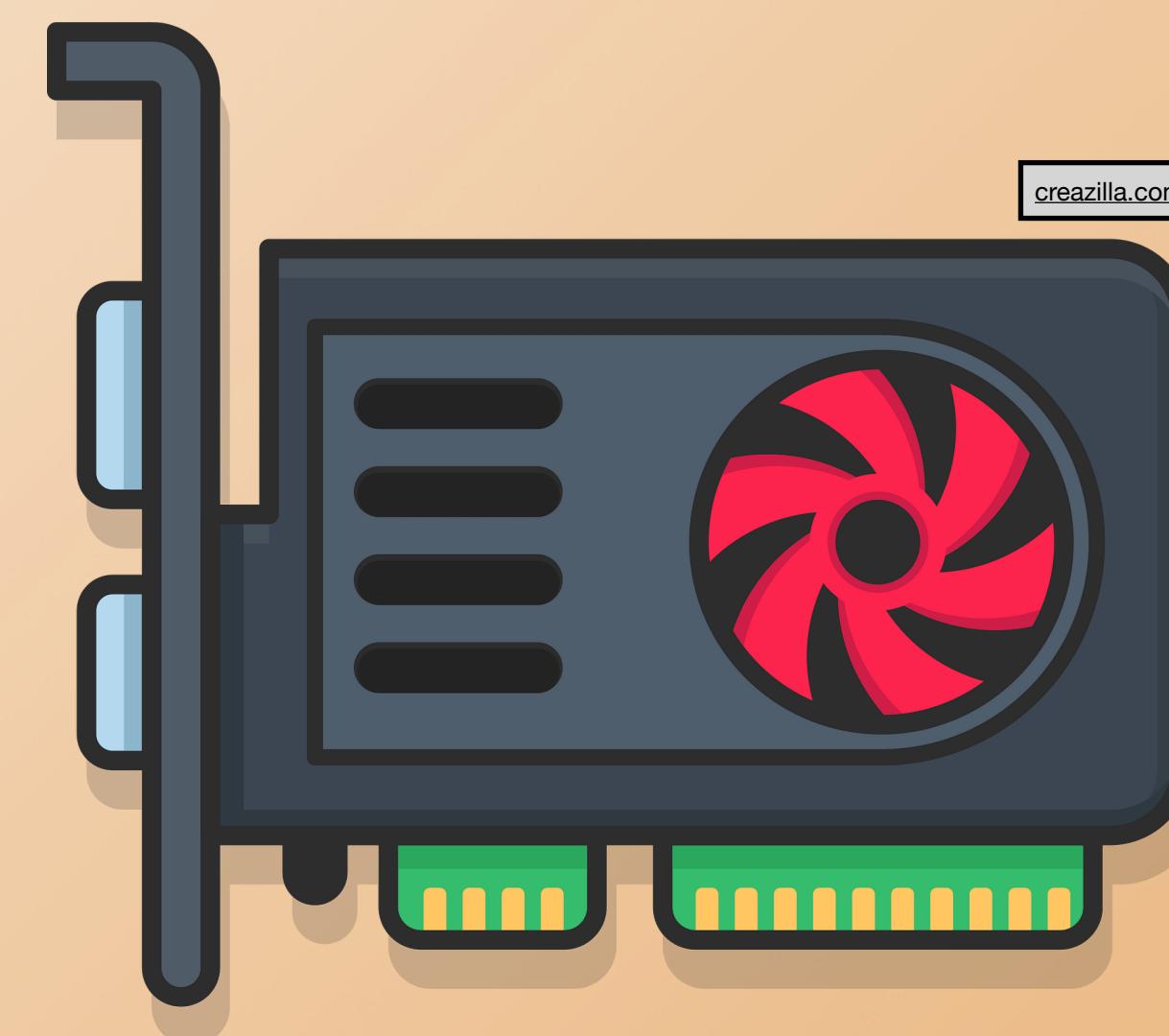
pl.pngtree.com

Utworzenie modelu do klasyfikacji pacjentów z podejrzeniem rozwoju nowotworu mózgu na podstawie zdjęć MRI

Parametry używanych maszyn



CPU: AMD Ryzen 7 2700 8/16
3.2 GHz
RAM: 32 GB GDDR4 3200 MHz



CPU: AMD Ryzen 7 2700 8/16 3.2 GHz
GPU: NVIDIA GTX 1660 1408 CUDA
1830 MHz
GPU vRAM: 6 GB GDDR5
RAM: 32GB GDDR4 3200 MHz

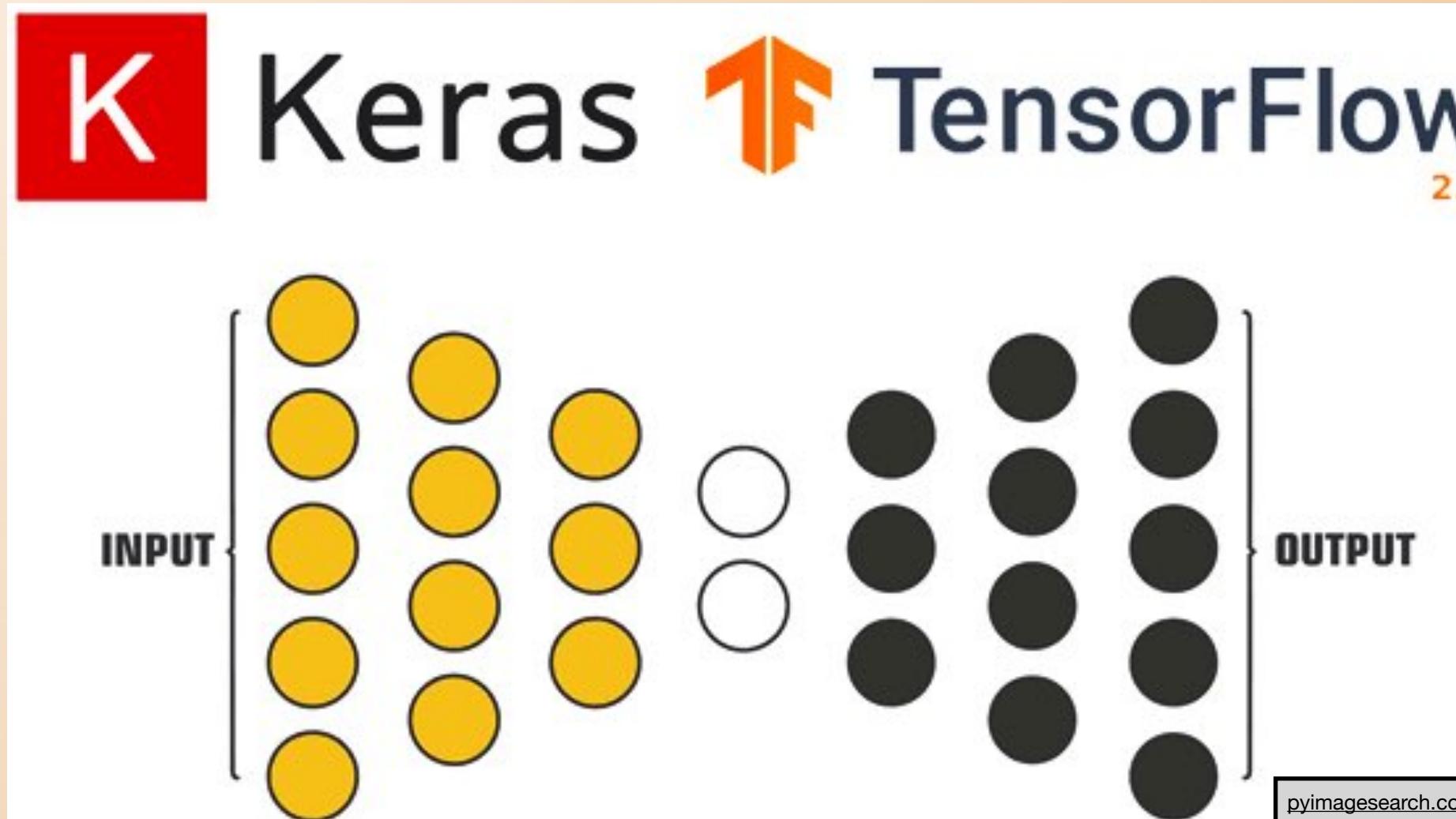


CPU: Intel Xeon 6/12 2.0 GHz
GPU: NVIDIA A100-SXM4 6912 CUDA
432 RT 765 MHz
GPU vRAM: 40 GB HMB2
RAM: 85 GB

Sposób działania

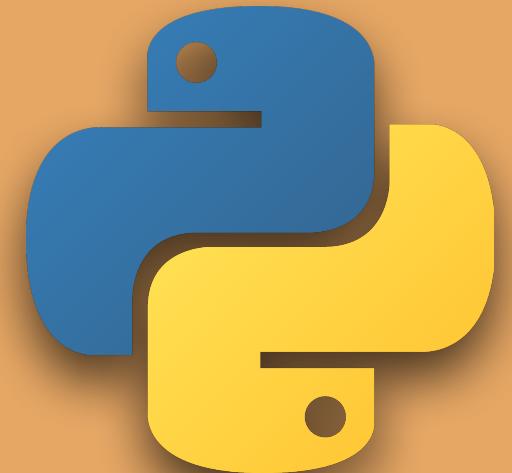
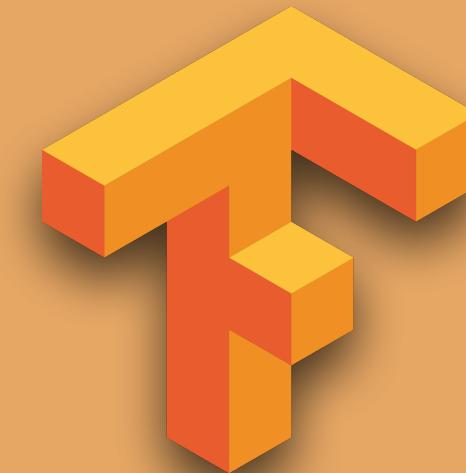
1. Import bibliotek i modułów.
2. Przygotowanie danych.
3. Podział naszych danych na zbiory testujący i uczący.
4. Pobranie modelu EfficientNetV2B0.
5. Odpowiednie ustawienie pobranego modelu.
6. Kompilacja modelu.
7. Odpowiednie ustawienie funkcji redukcji „*learning rate*”.
8. Trenowanie modelu.
9. Dokonanie predykcji.

Sposób działania



Przykładowe używane przez nas **funkcje** z paczki **Keras** z pakietu TensorFlow:

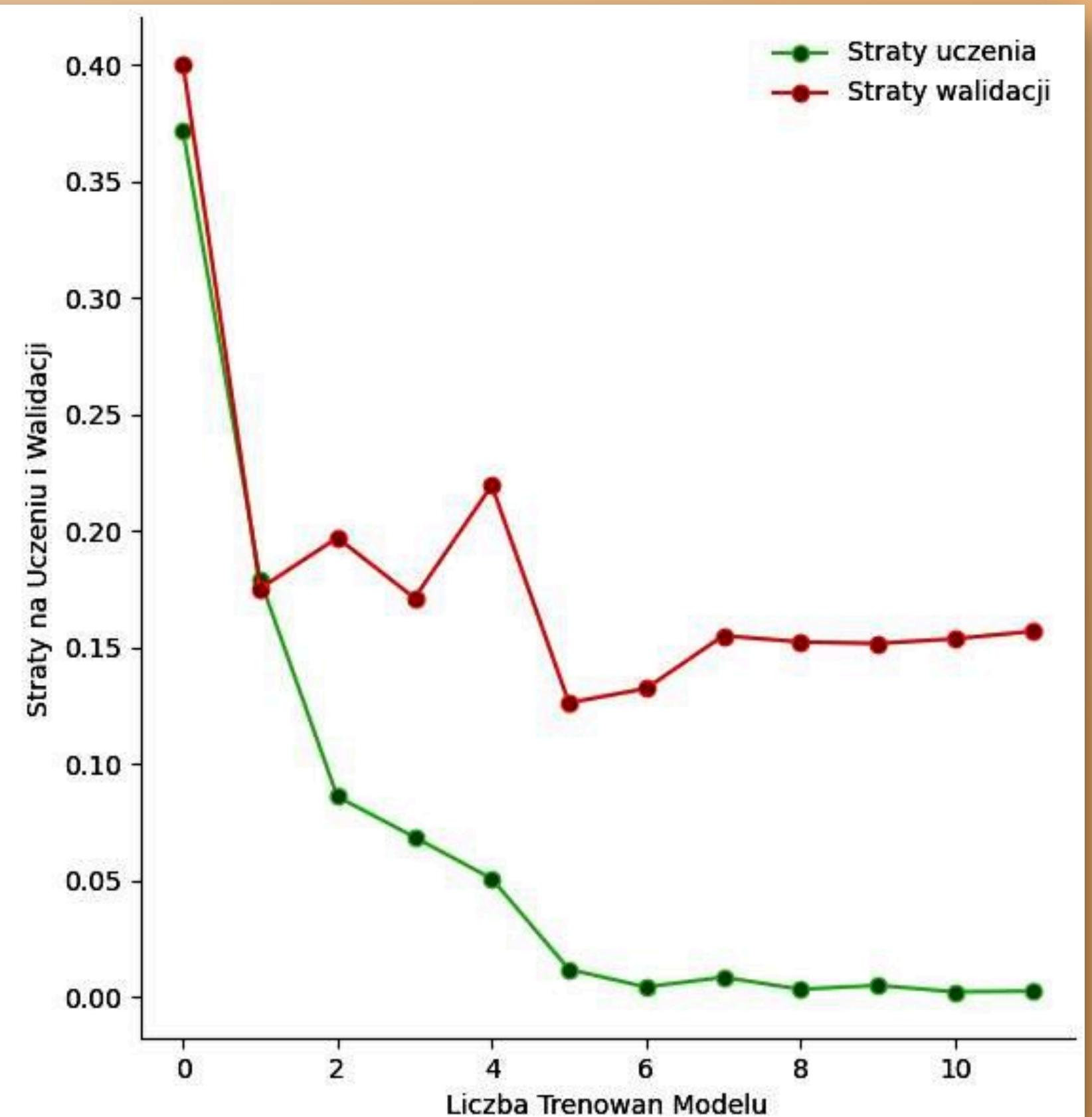
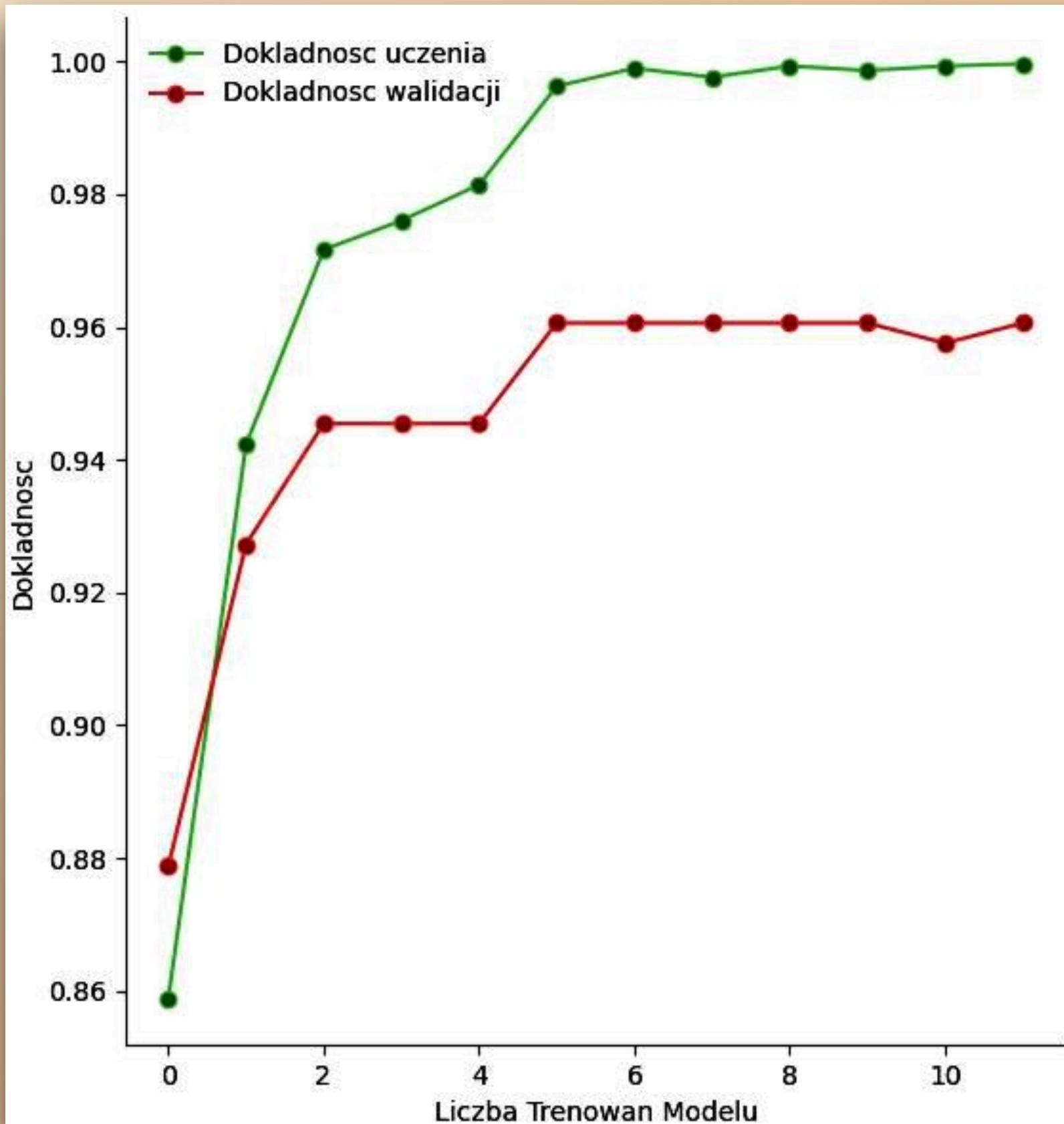
- ▶ *tf.keras.applications.EfficientNetV2B0,*
- ▶ *tf.keras.callbacks.TensorBoard,*
- ▶ *tf.keras.callbacks.ModelCheckpoint,*
- ▶ *tf.keras.callbacks.ReduceLROnPlateau,*
- ▶ *tf.keras.Model.fit().*



Prezentacja kodu

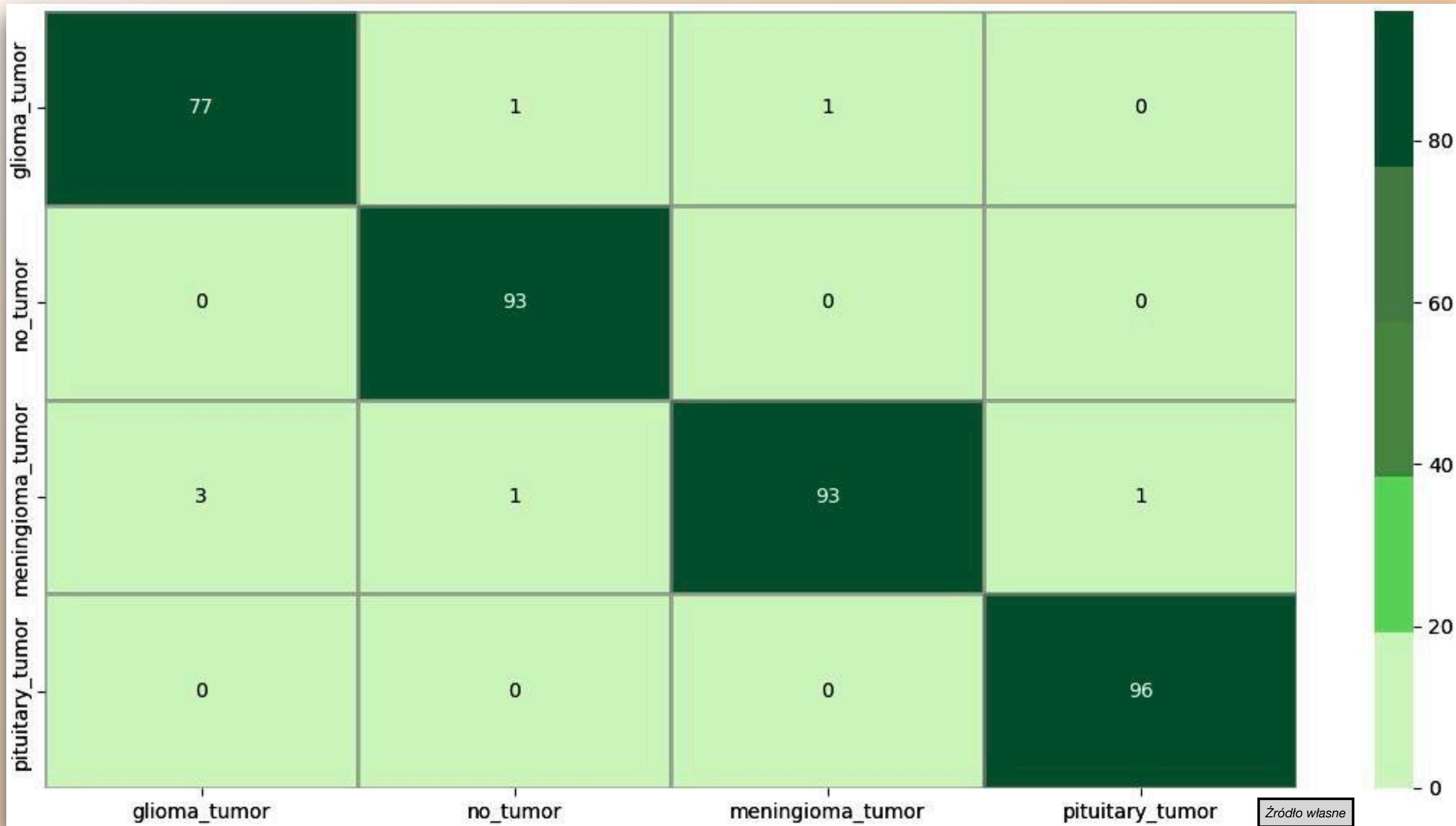
Wyniki

- Wraz ze wzrostem liczby trenowań modelu wzrasta poziom dokładności dla procesu uczenia oraz procesu walidacji.
- Wraz ze zwiększeniem liczby trenowań modelu znacząco spada liczba pomyłek w obu zbiorach.



Źródło własne

Wyniki



- ▶ Mapa cieplna przedstawia dokładność klasyfikacji obrazów MRI w odniesieniu do mocy modelu.
- ▶ Największą szansę wykrycia ma nowotwór przysadki, zaś najmniejszą glejak.

Wnioski

- ▶ Za pomocą **paczki TensorFlow** udało się przygotować **model rozpoznający odpowiedni rodzaj nowotworu mózgu.**
- ▶ Wraz ze wzrostem liczby trenowań naszego modelu, jego **dokładność dla zbiorów uczącego oraz testowego wzrastała**, natomiast **liczba pomyłek spadała**.
- ▶ **Mapa cieplna modelu** również potwierdziła jego **skuteczność**.
- ▶ **Optymalny model** uzyskujemy już przy 5. powtórzeniu.

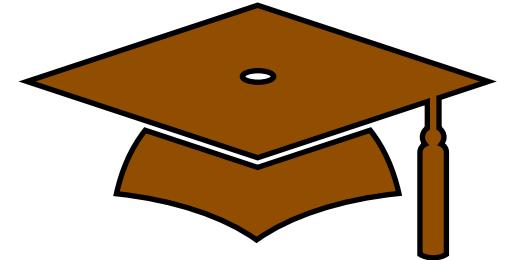
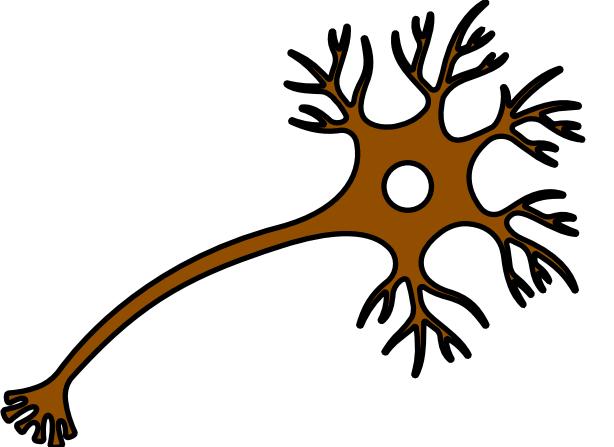
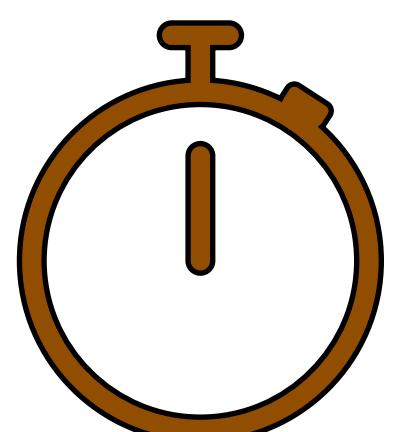
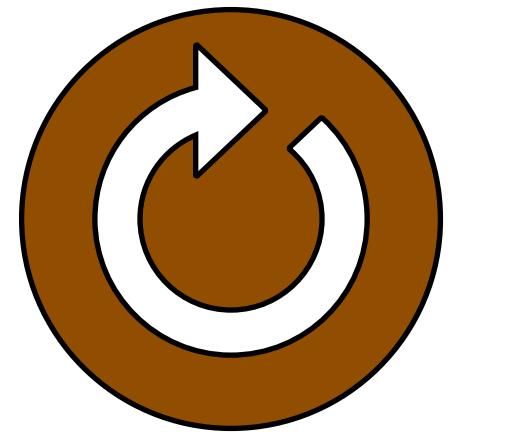
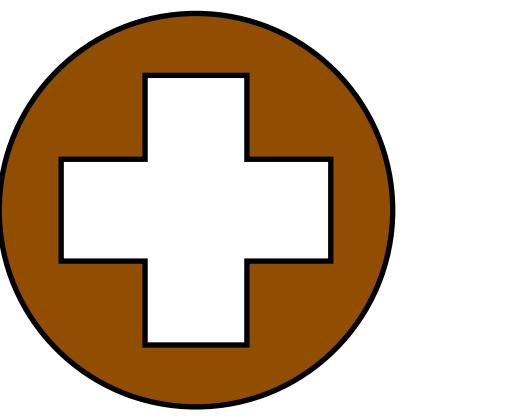


Inne możliwości



- ▶ Odpowiednikiem użytej biblioteki TensorFlow w środowisku Python jest **PyTorch**.
- ▶ Inną możliwością było zastosowanie **Scikit-learn**.

Podsumowanie

- ✓ Moduł Keras zaimplementowany w **pakiecie TensorFlow** służy do **uczenia maszynowego**, dzięki czemu jesteśmy w stanie stworzyć swoją sieć neuronową. 
-  ✓ **Splotowa sieć neuronowa (CNN)** wykorzystywana jest z powodzeniem m.in. w **rozpoznawaniu obrazów**, co ma zastosowanie w szerokiej gamie dziedzin badawczych.
- ✓ Celem uzyskania **najlepszego modelu** należy wykonać **większą liczbę powtórzeń przrównań**, uważając aby nie przetrenować otrzymanego modelu. 
-  ✓ **Wykorzystanie karty graficznej** wielokrotnie przyspiesza proces otrzymania właściwych modeli.
- ✓ Odpowiednio **wysoki stopień predykcji** jest niezwykle ważny w medycynie, w której to **niedopuszczalne** jest pominiecie właściwej klasyfikacji osoby chorej. 

Projekt przygotowali:

*Daria Plewa
Michał Humiński
Dominik Lisiecki*
Bioinformatyka

studia licencjackie rok 3 2021/2022



Seminarium bioinformatyczne

