

# Raport z projektu

Tomasz Mazur, Hubert Zięba

## Wstęp

Celem projektu było wytrenowanie modelu uczenia maszynowego do przewidywania wieku osób w oparciu o zdjęcia ich twarzy. Dane użyte do nauki i testowania można znaleźć tutaj. W ramach projektu rozważyliśmy następujące podproblemy:

1. Rozróżnić zdjęcia osób o wieku 2 i 50 lat
2. Rozróżnić zdjęcia osób o wieku 23 i 34 lata
3. Rozróżnianie 8 przedziałów wiekowych:
  - Od 1 do 3 lat
  - Od 4 do 8 lat
  - Od 9 do 15 lat
  - Od 16 do 25 lat
  - Od 26 do 40 lat
  - Od 41 do 60 lat
  - Od 61 do 80 lat
  - Powyżej 80 lat
4. Dla zdjęcia osoby dokładnie ocenić wiek (za pomocą klasyfikatora)
5. Dla zdjęcia osoby oszacować wiek (za pomocą modelu regresyjnego)

Dane zostały uprzednio podzielone na zbiór treningowy i testowy w stosunku 2 : 1. Dodatkowo rozpatrzyliśmy kilka różnych metod wstępnego przetworzenia danych:

- Brak przetwarzania
- Przetworzone za pomocą PCA w oparciu o zbiór treningowy
- Przetworzone za pomocą PCA z jądrem gaussowskim z parametrami  $\tau = 500, 1000$  i  $2000$  w oparciu o 10% zbioru treningowego (ze względów technicznych)

Do podzadań 1-4 użyliśmy SVM, jądrowego SVM, sieci neuronowych i konwolucyjnych sieci neuronowych, natomiast do podzadania 5 zostały wykorzystane odpowiednio zmodyfikowane sieci neuronowe i konwolucyjne sieci neuronowe oraz regresja grzbietowa z funkcjami jądrowymi.

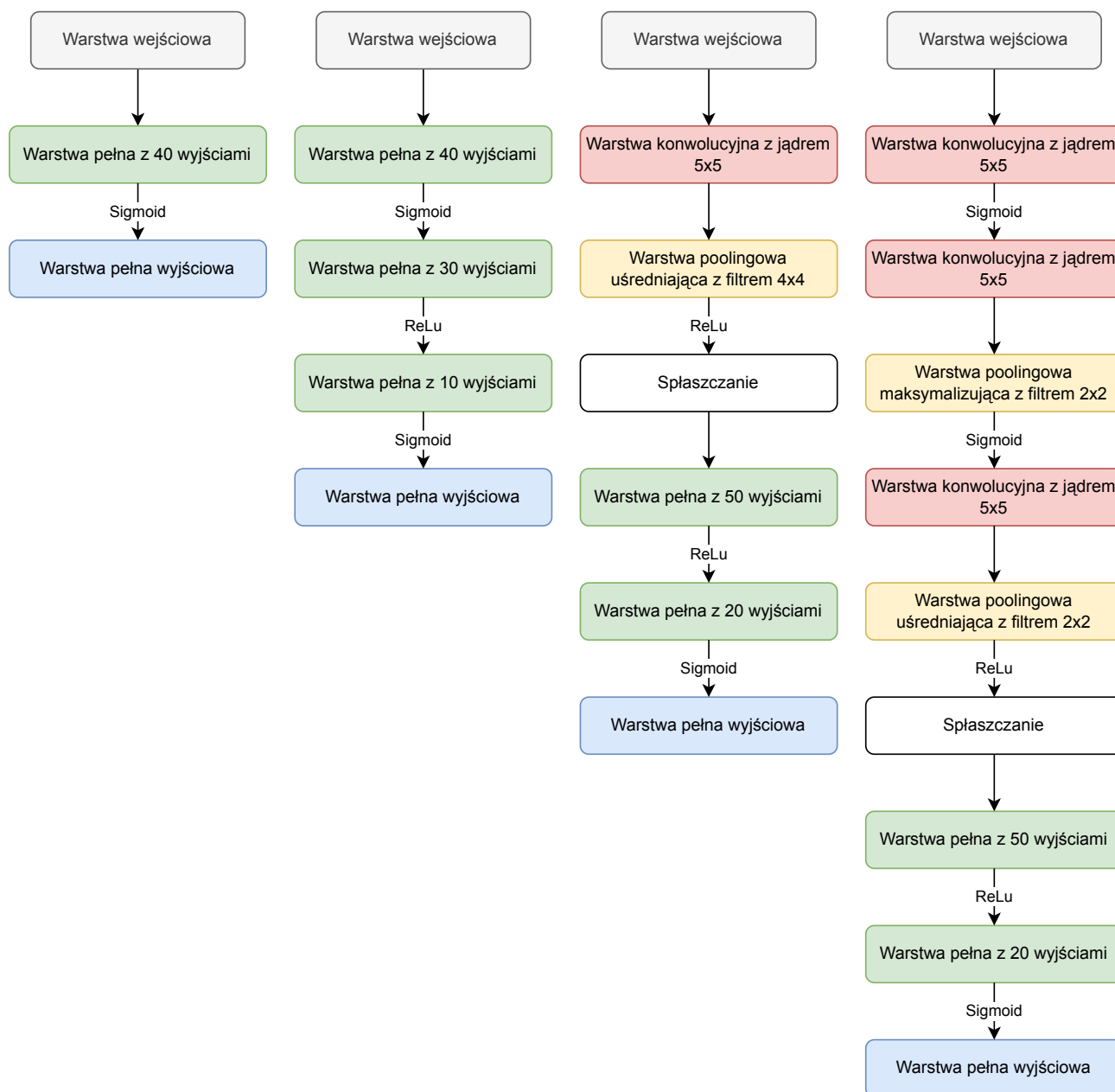
## Pojęcia

Do porównywania modeli zostały zastosowane dwie miary:

1. Dokładność –  $\frac{1}{n} \sum_{i=1}^n 1[h(x_i) = y_i]$ , gdzie  $x_i$  to dane wejściowe, a  $y_i$  to odpowiadające im dane wyjściowe. Jest to miara stosowana przy klasyfikacji, i jest to 1 - błąd klasyfikacji.
2. Średni błąd –  $\frac{1}{n} \sum_{i=1}^n |h(x_i) - y_i|$ . Ta miara była stosowana przy modelach, które szacowały wiek.

## Omówienie sieci neuronowych

Rozpatrzyliśmy 2 modele sieci neuronowych i 2 modele konwolucyjnych sieci neuronowych z odpowiednio zmodyfikowanymi warstwami wejściowymi i wyjściowymi. Parametry modeli zostały wybrane jako parametry najlepszego modelu z 10 losowych inicjalizacji. Proces nauki był złożony z 50 000 iteracji wstecznej propagacji z wykorzystaniem metody spadku wzdłuż gradientu. Struktura sieci znajduje się poniżej.



Rysunek 1: Struktura użytych sieci neuronowych. W dalszej części będziemy je oznaczać jako odpowiednio  $NN_1$ ,  $NN_2$ ,  $CNN_1$  i  $CNN_2$

## Omówienie maszyn wektorów nośnych

Rozpatrzyliśmy 2 modele oparte na maszynach wektorów nośnych. Jeden był oparty na jądrze gaussowskim, drugi na liniowej maszynie wektorów nośnych.

W obu przypadkach należało wzbogacić model SVM, aby móc klasyfikować wieloklasowo. Wybrany sposób na to był stworzenie prostego drzewa decyzyjnego, w którym każdy wierzchołek odpowiadał pewnemu przedziałowi klas  $[l, r)$ . W każdym wierzchołku drzewa była maszyna wektorów nośnych, która decydowała o tym, czy nadchodzące dane należą do przedziału  $[l, (l + r)/2)$ , czy do  $[(l + r)/2, r)$ . Takie podejście wymagało, aby istniał pewien naturalny porządek na klasach. Jako, że w naszych rozważaniach klasy odpowiadają przedziałom wieków, to istnieje naturalny porządek na tych klasach, w którym przedział niższych wieków ma niższy numer klasy, niż przedział starszych wieków.

Implementacje maszyn wektorów nośnych są wzorowane na pracy *Platt, John. (1998). Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines. Advances in Kernel Methods-Support Vector Learning. 208.*

W poniższych rozważaniach będą rozpatrywane jedynie modele gaussowskie z  $\gamma \in \{0.1, 3\}$ , gdzie  $\gamma = 1/(2\tau^2)$ , gdyż takie wartości osiągały najlepsze wyniki (i podobne do wartości pomiędzy nimi), a wszystkie modele wybierały na podstawie zbioru walidacyjnego najlepszy parametr ograniczający mnożniki Lagrange'a spośród możliwych wartości będących w zbiorze  $\{2^i : i \in \{-2, -1, \dots, 6\}\}$ .

Znajdujące poniżej wyniki dla modeli opartych na SVM są uśrednione na podstawie 3 lub 5 przebiegów na danych podzielonych losowo na zbiór treningowy, walidacyjny i testowy, przy czym dane były standaryzowane. Warto zwrócić uwagę, że modele oparte na SVM były najwolniejszymi spośród rozpatrzonych modeli, więc dlatego liczba przebiegów oraz czasem rozmiary zbiorów treningowych były mniejsze, niż u innych modeli. W szczególności nie byliśmy w stanie uzyskać wyników dla klasyfikacji wieloklasowej za pomocą modeli opartych na SVM na danych, które zostały przetworzone za pomocą jądrowego PCA z  $\tau \in \{500, 1000\}$ .

## Omówienie regresji grzbietowej

Najprostsze z rozpatrywanych modeli stosowały regresję grzbietową. Jeden model korzystał z klasycznej regresji grzbietowej, natomiast drugi korzystał z wersji z funkcjami gaussowskimi. Dla obu modeli rozwiązanie było liczone analitycznie, a parametr regularyzacji był wybierany na podstawie zbioru walidacyjnego. Ponadto w obu modelach do macierzy planowania został dodany stały współczynnik (kolumna jedynek).

## Rozróżnianie 2 i 50 lat

Ten zestaw służył za podstawowy test dla rozpatrywanych modeli, gdyż teoretycznie powinny one osiągać tutaj bardzo dobre wyniki. Wyniki prezentują się następująco:

Nazwa modelu	Sposób przetwarzania danych	Dokładność
Sieć $NN_1$	PCA	80.952%
Sieć $NN_2$	PCA	85.034%
Liniowy SVM	PCA	81.954%
SVM z $\gamma = 0.1$	PCA	76.322%
SVM z $\gamma = 3$	PCA	88.046%
Sieć $NN_1$	Jądrowe PCA z $\tau = 500$	55.102%
Sieć $NN_2$	Jądrowe PCA z $\tau = 500$	55.782%
Liniowy SVM	Jądrowe PCA z $\tau = 500$	56.896%
SVM z $\gamma = 0.1$	Jądrowe PCA z $\tau = 500$	55.402%
SVM z $\gamma = 3$	Jądrowe PCA z $\tau = 500$	55.517%
Sieć $NN_1$	Jądrowe PCA z $\tau = 1000$	89.456%
Sieć $NN_2$	Jądrowe PCA z $\tau = 1000$	87.755%
Liniowy SVM	Jądrowe PCA z $\tau = 1000$	55.748%
SVM z $\gamma = 0.1$	Jądrowe PCA z $\tau = 1000$	55.747%
SVM z $\gamma = 3$	Jądrowe PCA z $\tau = 1000$	55.747%
Sieć $NN_1$	Jądrowe PCA z $\tau = 2000$	81.633%
Sieć $NN_2$	Jądrowe PCA z $\tau = 2000$	84.694%
Liniowy SVM	Jądrowe PCA z $\tau = 2000$	76.436%
SVM z $\gamma = 0.1$	Jądrowe PCA z $\tau = 2000$	55.747%
SVM z $\gamma = 3$	Jądrowe PCA z $\tau = 2000$	55.747%
Sieć $CNN_1$	Brak	60.204%
Sieć $CNN_2$	Brak	55.782%

Niektóre modele nie spełniły oczekiwań, jak np. jądrowy SVM z jądrowym PCA czy konwolucyjne sieci. Inne skolei osiągały dokładność na poziomie 90%. Co więcej, podczas testów lokalnych zdarzało się, że przy niektórych konfiguracjach sieci neuronowych modele notowały dokładność rzędu 95%, niestety przy późniejszych próbach wyniki wychodziły na ogół gorsze.

## Rozróżnianie 23 i 34 lat

Ten zestaw miał być natomiast jednym z trudniejszych i nie oczekiwaliśmy od modeli świetnych rezultatów. Natomiast pojedyncze algorytmy osiągały nawet 66% skuteczności.

Nazwa modelu	Sposób przetwarzania danych	Dokładność
Sieć $NN_1$	PCA	58.947%
Sieć $NN_2$	PCA	58.597%
Liniowy SVM	PCA	53.333%
SVM z $\gamma = 0.1$	PCA	60.119%
SVM z $\gamma = 3$	PCA	61.310%
Sieć $NN_1$	Jądrove PCA z $\tau = 500$	55.088%
Sieć $NN_2$	Jądrove PCA z $\tau = 500$	51.228%
Liniowy SVM	Jądrove PCA z $\tau = 500$	53.571%
SVM z $\gamma = 0.1$	Jądrove PCA z $\tau = 500$	56.904%
SVM z $\gamma = 3$	Jądrove PCA z $\tau = 500$	53.928%
Sieć $NN_1$	Jądrove PCA z $\tau = 1000$	60.702%
Sieć $NN_2$	Jądrove PCA z $\tau = 1000$	56.140%
Liniowy SVM	Jądrove PCA z $\tau = 1000$	51.190%
SVM z $\gamma = 0.1$	Jądrove PCA z $\tau = 1000$	51.191%
SVM z $\gamma = 3$	Jądrove PCA z $\tau = 1000$	51.191%
Sieć $NN_1$	Jądrove PCA z $\tau = 2000$	65.965%
Sieć $NN_2$	Jądrove PCA z $\tau = 2000$	61.754%
Liniowy SVM	Jądrove PCA z $\tau = 2000$	58.214%
SVM z $\gamma = 0.1$	Jądrove PCA z $\tau = 2000$	51.190%
SVM z $\gamma = 3$	Jądrove PCA z $\tau = 2000$	51.190%
Sieć $CNN_1$	Brak	65.306%
Sieć $CNN_2$	Brak	59.524%

## Rozróżnianie przedziałów wiekowych

W celu przetestowania różnych klasyfikatorów podzieliliśmy arbitralnie dane na przedziały, jako że osoby w podobnym wieku będą do siebie podobne. W tym przypadku wszystkie modele odnotowywały wyniki na poziomie 40%. Przy 8 klasach ten wynik można uznać za dobry.

Nazwa modelu	Sposób przetwarzania danych	Dokładność
Sieć $NN_1$	PCA	40.328%
Sieć $NN_2$	PCA	42.091%
Liniowy SVM	PCA	39.486%
SVM z $\gamma = 0.1$	PCA	39.487%
SVM z $\gamma = 3$	PCA	40.570%
Sieć $NN_1$	Jądrowe PCA z $\tau = 500$	39.342%
Sieć $NN_2$	Jądrowe PCA z $\tau = 500$	39.342%
Sieć $NN_1$	Jądrowe PCA z $\tau = 1000$	39.490%
Sieć $NN_2$	Jądrowe PCA z $\tau = 1000$	39.342%
Sieć $NN_1$	Jądrowe PCA z $\tau = 2000$	40.624%
Sieć $NN_2$	Jądrowe PCA z $\tau = 2000$	39.342%
Liniowy SVM	Jądrowe PCA z $\tau = 2000$	39.647%
SVM z $\gamma = 0.1$	Jądrowe PCA z $\tau = 2000$	39.488%
SVM z $\gamma = 3$	Jądrowe PCA z $\tau = 2000$	39.488%
Sieć $CNN_1$	Brak	39.342%
Sieć $CNN_2$	Brak	39.342%

## Dokładne ocenianie wieku

Rozważyliśmy również dość ekstremalny przypadek, w którym każdy wiek tworzy jedną klasę. W tym podproblemie duże znaczenie miał fakt, że dane nie były rozłożone równomiernie, więc większość modeli skupiała się na poprawnym klasyfikowaniu zdjęć osób w wieku 20-30 lat, co też skutkuje tym, że większość modeli uzyskiwała wynik w okolicach 9%. Warto wyróżnić sieć  $NN_1$  w połączeniu z PCA za osiągnięcie aż 11% poprawności zatem można powiedzieć, że jej udało się dostrzec realne zależności między danymi, a nie tylko ich rozkładem.

Nazwa modelu	Sposób przetwarzania danych	Dokładność
Sieć $NN_1$	PCA	11.281%
Sieć $NN_2$	PCA	9.210%
Liniowy SVM	PCA	9.270%
SVM z $\gamma = 0.1$	PCA	9.252%
SVM z $\gamma = 3$	PCA	5.898%
Sieć $NN_1$	Jądrowe PCA z $\tau = 500$	9.197%
Sieć $NN_2$	Jądrowe PCA z $\tau = 500$	9.210%
Sieć $NN_1$	Jądrowe PCA z $\tau = 1000$	9.173%
Sieć $NN_2$	Jądrowe PCA z $\tau = 1000$	9.210%
Sieć $NN_1$	Jądrowe PCA z $\tau = 2000$	9.629%
Sieć $NN_2$	Jądrowe PCA z $\tau = 2000$	9.210%
Liniowy SVM	Jądrowe PCA z $\tau = 2000$	9.408%
SVM z $\gamma = 0.1$	Jądrowe PCA z $\tau = 2000$	9.252%
SVM z $\gamma = 3$	Jądrowe PCA z $\tau = 2000$	9.252%
Sieć $CNN_1$	Brak	9.210%
Sieć $CNN_2$	Brak	9.210%

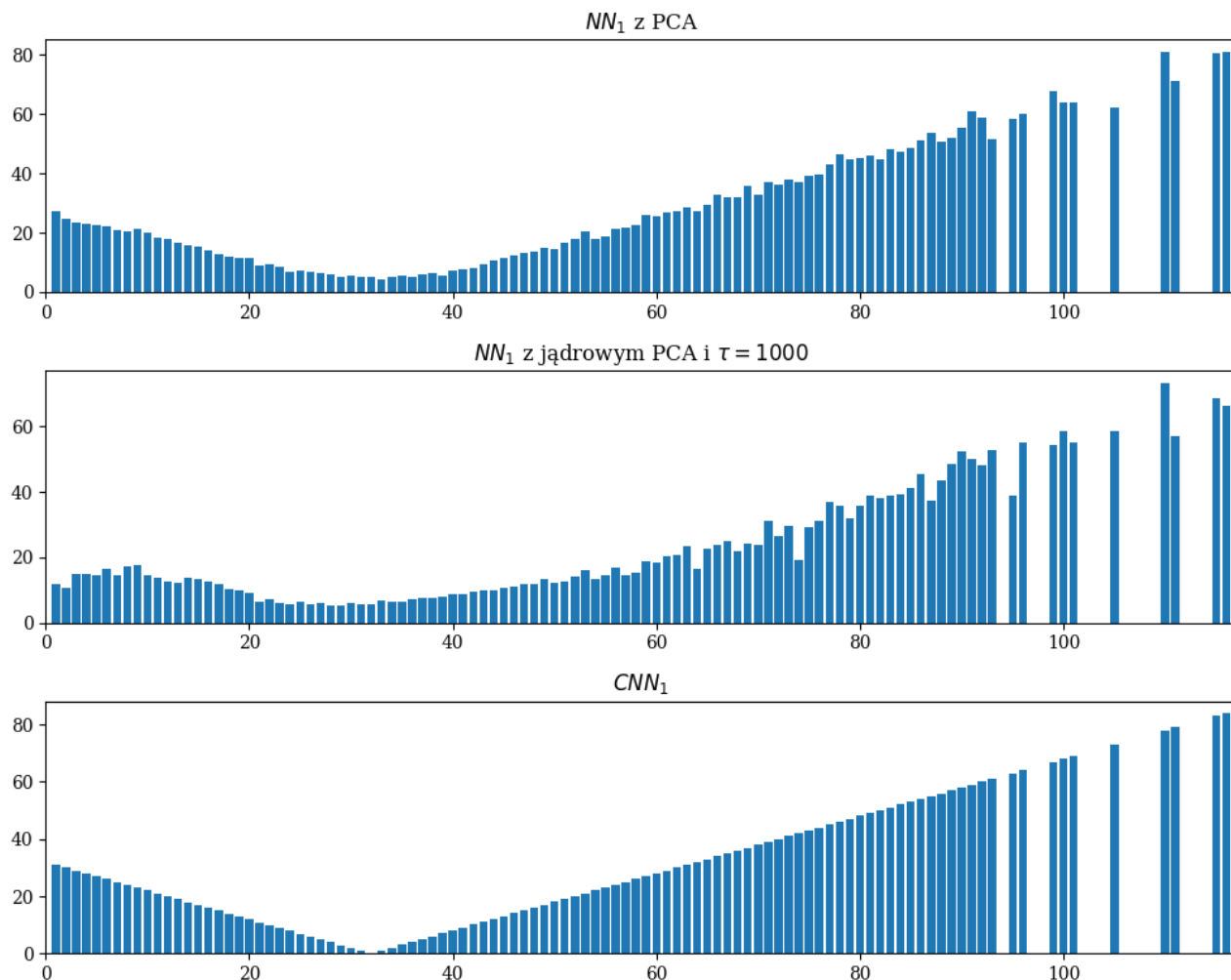


## Szacowanie wieku

W ostatnim podzadaniu zamiast podchodzić do naszego problemu jako problemu klasyfikacji spróbowa-  
liśmy rozważyć modele regresyjne. Najlepsze wyniki osiągnęła sieć  $NN_1$  w połączeniu z jądrowym PCA  
myląc się średnio o 11 lat.

Nazwa modelu	Sposób przetwarzania danych	Średni błąd
Sieć $NN_1$	PCA	14.380
Sieć $NN_2$	PCA	15.240
Regresja liniowa	PCA	12.664
Sieć $NN_1$	Jądrowe PCA z $\tau = 500$	15.239
Sieć $NN_2$	Jądrowe PCA z $\tau = 500$	15.240
Sieć $NN_1$	Jądrowe PCA z $\tau = 1000$	11.389
Sieć $NN_2$	Jądrowe PCA z $\tau = 1000$	15.240
Sieć $NN_1$	Jądrowe PCA z $\tau = 2000$	13.611
Sieć $NN_2$	Jądrowe PCA z $\tau = 2000$	15.240
Regresja liniowa	Jądrowe PCA z $\tau = 2000$	14.172
Regresja z $\gamma = 1$	Jądrowe PCA z $\tau = 2000$	15.451
Sieć $CNN_1$	Brak	15.226
Sieć $CNN_2$	Brak	15.146

Dodatkowo dla wybranych wykresów zrobiliśmy analizę w jak dla osób w poszczególnych wiekach modele się myliły. Wyniki są dość zaskakujące, gdyż większość modeli miała tendencję do oceniania osób na zdjęciach na 20-25 lat. Tak jak poprzednio, w dużym stopniu może to wynikać z bardzo nierównomiernego rozkładu danych.



## Wnioski i spostrzeżenia

Niestety otrzymane wyniki nie były w pełni satysfakcjonujące. Ogólnie można powiedzieć, że proste sieci neuronowe sprawdzały się najlepiej, a drugą najlepszą opcją była regresja grzbietowa, która nad innymi modelami zyskiwała zarówno na wynikach, jak i prostocie implementacji oraz wymaganiach obliczeniowych.

Analiza składowych głównych w dużym stopniu upraszczała obliczenia, nie pogarszając znacząco wyników (zwykle PCA redukowało liczbę cech z 2304 do 172). Nietkóre modele (SVM i regresja grzbietowa) potrzebowały operować na danych przerobionych za pomocą PCA, by otrzymać jakiegokolwiek wyniki.

Konwolucyjne sieci neuronowe liczyły się istotnie dłużej od pozostałych modeli (kilka do kilkunastu minut na jeden przebieg na jednym zbiorze treningowym) i nie dawały oczekiwanych rezultatów. Co warto zaznaczyć w przypadku sieci, przy niektórych strukturach, jak i odpowiednim dobraniu parametrów do metody spadku wzdłuż gradientu, modele otrzymywały lepsze rezultaty, niestety było to na tyle kruche, że nie udało się ich zrekonstruować na etapie generowania danych. Pokazuje to jak bardzo sieci neuronowe są ciężkie do skonfigurowania, natomiast skutkuje to czasem ponadprzeciętnymi wynikami.

W przypadku SVM jedną z większych przeszkód były wymagania obliczeniowe, przez które model musiał być trenowany z mniejszą dokładnością oraz na mniejszych zbiorach treningowych, ze sprawdzeniem mniejszej liczby hiperparametrów, od których w dużej mierze zależy dokładność modelu.