

Wyraźny wzrost ilości generowanych danych (o różnej jakości) jest widoczny w wielu dziedzinach, takich jak obrazowanie medyczne, kategoryzacja tekstów, biologia obliczeniowa, genomika czy bankowość. Mogłoby się wydawać, że wzrost ten jest bardzo korzystny – więcej danych może oznaczać więcej możliwości wyekstrahowania istotnych informacji i wiedzy dotyczących analizowanego problemu. W wielu sytuacjach mamy jednak do czynienia z sytuacją odwrotną – efektywne przetwarzanie i analiza dużych oraz „trudnych” danych stanowi kluczowy problem, z którym mierzą się społeczności naukowe, zwłaszcza w erze „dużych danych” (ang. *big data*). Rewolucja w zakresie takich zbiorów danych wyraźnie wpłynęła na wiele dziedzin nauki i przemysłu – statystykę, uczenie maszynowe, obliczenia równoległe, czy systemy komputerowe i obliczeniowe w ogólności. W ramach naszego projektu zajmiemy się przede wszystkim analizą danych wizji komputerowej – obrazami oraz sekwencjami wideo, jednak opracowane metody będą mogły znaleźć zastosowanie do przetwarzania także innych rodzajów danych.

W klasycznym podejściu analiza obrazów składa się z dwóch etapów: ekstrakcji cech oraz ich klasyfikacji. Pierwszy etap polega na ekstrakowaniu pewnych charakterystycznych cech, dzięki czemu obraz może zostać przedstawiony w postaci ciągu liczb opisujących już nie sam obraz, ale obiekt widoczny w obrazie. Przykładowo dla obrazu twarzy takimi cechami mogłaby być długość nosa względem rozstawu oczu, czy stopień zaokrąglenia podbródka. Cechy te muszą być „ręcznie” definiowane przez inżynierów i programistów, którzy projektują odpowiednie algorytmy służące do analizy obiektów określonego rodzaju. Następnie wyekstrahowane wektory cech są klasyfikowane, czyli przypisywane do określonej grupy obiektów (na przykład może to oznaczać podjęcie decyzji, czy osoba widoczna na zdjęciu ma ubrane okulary). Stosowane są tutaj tzw. *klasyfikatory*, które uczą się reguł klasyfikacji na podstawie zbiorów zawierających odpowiednią liczbę przykładów (kontynuując wcześniejszy przykład, klasyfikator mógłby zostać wytrenowany na podstawie kilkuset zdjęć przedstawiających osoby w okularach i kilkuset zdjęć osób bez okularów). Za jeden z najbardziej znanych i najlepiej przetestowanych klasyfikatorów można uznać *maszynę wektorów podpierających* (ang. *support vector machine* – SVM). Znalazła ona zastosowanie w rozwiązywaniu wielu zadań związanych z rozpoznawaniem wzorców i stanowiła do niedawna naturalny wybór przy projektowaniu nowych systemów klasyfikacyjnych.

Obecnie jesteśmy świadkami rewolucji w świecie wizji komputerowej oraz uczenia maszynowego, polegającej na pojawieniu się i popularyzacji algorytmów *uczenia głębokiego*. Jest to zbiorcza nazwa całej grupy metod wykorzystujących wielowarstwowe sieci neuronowe, które pozwalają na naukę zarówno cech, które są ekstrahowane, jak i reguł klasyfikacyjnych. W celu analizy obrazów oraz sekwencji wideo opracowane zostały głębokie sieci konwolucyjne – początkowe warstwy sieci realizują ekstrakcję cech za pomocą filtrów konwolucyjnych, a warstwy końcowe dokonują klasyfikacji wyekstrahowanych cech. Dzięki temu można uniknąć żmudnego procesu projektowania i doboru cech, które muszą zostać z obrazu wyekstrahowane. Z drugiej strony pojawił się inny problem – architektura sieci konwolucyjnej ma ogromny wpływ na uzyskiwane wyniki, a efektywne jej zaprojektowanie wymaga intuicji i dużego doświadczenia.

Potencjalnie do klasyfikacji cech wyuczonych przez sieć konwolucyjną można by było zastosować SVM, który jest uważany za bardziej skuteczny klasyfikator niż sieci neuronowe. Takie podejście nie jest jednak stosowane z uwagi na pewne ograniczenia SVM-a. Zaliczają się do nich czasowo i pamięciowo kosztowny trening, trudności w doborze parametrów (przede wszystkim tzw. *funkcji jądrowej*), czy trudności w doborze „przydatnych” cech oraz elementów zbioru treningowego. Problemy te są szczególnie istotne w przypadku sieci głębokich, które z kolei wymagają ogromnych ilości danych do tego, żeby mogły zostać poprawnie nauczone. Niniejszy projekt ma na celu pokonanie tych ograniczeń z pomocą algorytmów ewolucyjnych, pozwalających na znalezienie najlepszego (lub prawie najlepszego) rozwiązania poprzez zastosowanie mechanizmów zbliżonych do biologicznej ewolucji. W efekcie pozwoli to na zastosowanie SVM-a do klasyfikacji danych trudnych, do których zdecydowanie należy zaliczyć cechy ekstrahowane przez głębokie sieci konwolucyjne – naszym celem jest poprawa jakości działania tych ostatnich i obniżenie wpływu architektury sieci na uzyskiwane wyniki.

Realizacja projektu umożliwi połączenie zalet głębokich sieci konwolucyjnych z wysoką skutecznością klasyfikacji za pomocą SVM-a. Dzięki odpowiedniemu wykorzystaniu algorytmów ewolucyjnych duże ilości danych będą mogły przestać być przeszkodą w stosowaniu SVM-a, a wręcz zaczną stanowić zaletę, pozwalając na osiągnięcie lepszych wyników klasyfikacji. W trakcie projektu planujemy zweryfikować możliwości projektowanych rozwiązań w wielu zagadnieniach komputerowej wizji. Wśród nich znajdzie się rozpoznawanie twarzy oraz gestów, analiza stanów emocjonalnych na podstawie sekwencji wideo, a także analiza obrazów medycznych na potrzeby diagnostyki nowotworowej.