Studium Magisterskie

Kierunek metody ilościowe i systemy informacyjne w ekonomii

Jan Chrzanowski

Nr albumu 123169

Automatyczna ocena wieku na podstawie obrazów twarzy za pomocą konwolucyjnych sieci neuronowych

Praca Magisterska

Pod kierunkiem naukowym

Dr Jarosław Olejniczak

Zakład Technologii Informatycznych

Warszawa 2024

Spis treści

Wstęp

Współczesne społeczeństwo coraz częściej styka się z problemem związanym z identyfikacją wieku na podstawie zdjęć twarzy. Automatyczna ocena wieku osoby za pomocą technologii biometrycznych staje się coraz bardziej istotna w wielu dziedzinach życia, od marketingu i handlu detalicznego, przez medycynę, aż po różne aspekty bezpieczeństwa publicznego. Rozwój głębokich sieci neuronowych oferuje nowe możliwości w zakresie analizy i przetwarzania obrazów, co może znacząco poprawić dokładność i efektywność systemów do oceny wieku.

Automatyczna ocena wieku na podstawie zdjęć twarzy ma potencjalne zastosowanie w wielu obszarach. W handlu detalicznym, może być wykorzystywana do personalizacji doświadczeń zakupowych, dostosowywania ofert i rekomendacji produktów do wieku klientów. W medycynie, systemy te mogą wspierać diagnozowanie niektórych schorzeń, których symptomy mogą być związane z wiekiem pacjenta. W kontekście bezpieczeństwa, technologie te mogą być używane do monitorowania miejsc publicznych, kontrolowania dostępu do określonych stref oraz weryfikacji tożsamości oraz wieku na stronach internetowych. Jak podaje serwis IT biznes kamery automatycznie zweryfikują wiek klienta kupującego alkohol w sklepie. System wykorzystuje kamery z algorytmem, który ocenia wiek na podstawie twarzy klienta. Jeśli wiek klienta zostanie oszacowany poniżej 25 lat, konieczne będzie okazanie dowodu tożsamości. Technologia została przetestowana na 125 000 twarzach, osiągając średnią dokładność szacowania wieku z błędem około 2,2 roku. System ten zwiększa wygodę zakupów, eliminując konieczność angażowania personelu.

Jednak rozwój i wdrożenie takich systemów wiąże się z pewnymi wyzwaniami. Przede wszystkim, dokładność oceny wieku jest kluczowa, a jej osiągnięcie wymaga zaawansowanych algorytmów oraz wysokiej, jakości danych treningowych. Głębokie sieci neuronowe, ze względu na swoją zdolność do automatycznego wyodrębniania cech z surowych danych, są obiecującą technologią w tej dziedzinie. Konieczne jest jednak odpowiednie przygotowanie i przetworzenie danych, aby model mógł uczyć się na podstawie zróżnicowanych przypadków.

Ocena wieku za pomocą zdjęć twarzy rodzi również pytania etyczne i związane z prywatnością. Wykorzystanie biometrii do analizy twarzy może budzić obawy dotyczące naruszania prywatności oraz możliwości nadużyć. Istotne jest, aby systemy te były

projektowane i wdrażane w sposób, który zapewnia ochronę danych osobowych oraz zgodność z obowiązującymi przepisami prawnymi.

Celem niniejszej pracy jest stworzenie modelu głębokiej sieci neuronowej, który będzie w stanie ocenić, czy dana osoba jest pełnoletnia na podstawie zdjęcia twarzy. W pracy zostaną przedstawione dostępne metody i podejścia w zakresie oceny wieku, omówione zostaną różne zbiory danych wykorzystywane w badaniach oraz przeanalizowane będą etyczne aspekty związane z wykorzystaniem ludzkich zdjęć. Model zostanie przetestowany na dostępnych danych treningowych i testowych, a wyniki będą poddane szczegółowej analizie pod kątem dokładności i praktycznej przydatności.

Przegląd literatury i danych w pierwszym rozdziale pozwoli na zrozumienie aktualnego stanu badań w dziedzinie oceny wieku za pomocą technik uczenia maszynowego i głębokich sieci neuronowych. Omówione zostaną istniejące zbiory danych oraz najnowsze osiągnięcia w tej dziedzinie. Rozdział drugi skupi się na opisie oraz przeglądzie modeli, które mogą być zastosowane w analizie. W tym rozdziale zostanie również zwrócona uwaga na proces przygotowywania danych wejściowych w taki sposób, aby zmaksymalizować potencjalne osiągi modelu. W rozdziale trzecim zaprezentowane zostanie proces tworzenia modelu, jego trenowanie oraz ewaluacja na danych testowych, a także będą szczegółowo opisane wyniki oraz ocena skuteczności modelu, w tym analiza struktury modelu oraz jego ograniczeń.

Podsumowując, praca ta nie tylko ma na celu stworzenie efektywnego modelu do oceny wieku, ale również zwraca uwagę na potencjalne korzyści i wyzwania związane z implementacją tego typu systemów w rzeczywistych zastosowaniach. Przedstawienie możliwych kierunków rozwoju i zastosowań tej technologii może przyczynić się do dalszego postępu w dziedzinie analizy biometrycznej oraz do zrozumienia jej wpływu na różne aspekty życia społecznego i gospodarczego.

Rozdział I. Przegląd literatury i danych

1.1 Wprowadzenie do problemu oceny wieku po zdjęciu

Proces starzenia się twarzy jest wynikiem dynamicznego i kumulacyjnego wpływu czasu na skórę, tkanki miękkie oraz głębokie struktury twarzy. Charakteryzuje się skomplikowaną synergią zmian w teksturze skóry i utraty objętości twarzy. Wiele oznak starzenia wynika z kombinacji efektów grawitacji, postępującej resorpcji kości, zmniejszenia elastyczności tkanek oraz redystrybucji objętości podskórnej (Coleman i Grover, 2016, s. 4-6). Oznacza to, że nasza twarz bezpośrednio odzwierciedla nasz proces starzenia. W literaturze często możemy natrafić na szereg czynników, które wraz z upływem czasu wpływają na wygląd twarzy. W procesie oceny wieku na podstawie zdjęcia twarzy, warto jednak skupić się na konsekwencjach tych zmian takich jak, zmiana objętości twarzy, proporcje czy też zmiany skórne. Jednak sam proces oceny wieku na podstawie zdjęcia napotyka na szereg wyzwań. Głównym wyzwaniem jest to, że bazując tylko na zdjęciach rozróżniamy tak naprawdę wiek biologiczny osoby a nie jej wiek charakterystyczny, (czyli ten zgodny z datą urodzenia). To sprawia, że wiele czynników takich na przykład genetyka, odżywianie, kształt ciała, stan zdrowia, sprawność sercowo-oddechowa, warunki społeczne oraz styl życia, będą bezpośrednio wpływać na wygląd twarzy danej osoby, co może sprawić, że ich faktyczny wiek a ten oceniany na bazie zdjęcia może się różnić (Angulu i in., 2018, s. 1-2). Należy jednak zadać pytanie czy komputer może ocenić wiek na podstawie zdjęcia lepiej od człowieka? Naukowcy z Michigan State University przeprowadzili analizę, porównując wydajność estymacji wieku z obrazów twarzy przez ludzi oraz maszyny. W artykule zaproponowali oni hierarchiczne podejście do estymacji wieku, które składa się z przetwarzania wstępnego następnie lokalizację komponentów twarzy, w kolejnym kroku ekstrakcje cech a na koniec hierarchiczną estymację wieku. Wyniki porównano z oszacowaniami ludzi. Co ciekawe proponowane podejście hierarchiczne odniosło zwycięstwo, nad oszacowaniami według ludzi, tym samym prowadząc do dokładniejszych estymacji wieku (Han i in., 2013, s. 7). Jak widać problem ten może wydawać się być prostszy dla maszyny niż dla ludzkiego oka, jednakże nie należy tutaj od razu gloryfikować aktualne modele, ponieważ cały proces jest zależny od bardzo wielu czynników, które będą poruszane w dalszej części pracy.

Ocena wieku na podstawie zdjęcia twarzy to złożone zadanie, które napotyka na liczne wyzwania. Choć z pozoru wydaje się proste, istnieje wiele czynników, które mogą wpłynąć na dokładność estymacji przez modele oparte na głębokich sieciach neuronowych.

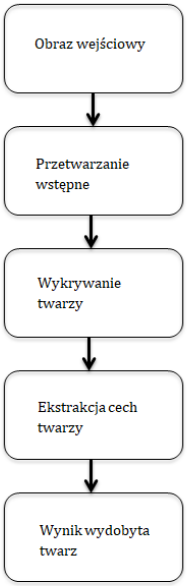
Jednym z kluczowych wyzwań jest zmienność w pozycjonowaniu głowy oraz wyrównaniu obrazu. Zdjęcia mogą być wykonane pod różnymi kątami, co zniekształca cechy twarzy i utrudnia poprawne oszacowanie wieku. Rozwiązaniem tego problemu jest stosowanie technik wyrównania twarzy, które ujednolicają pozycję twarzy na zdjęciach. Kolejnym problemem jest rozdzielczość obrazu. Niska rozdzielczość powoduje utratę istotnych szczegółów, takich jak zmarszczki czy tekstura skóry, co prowadzi do niedokładnych wyników. Z kolei bardzo wysoka rozdzielczość zwiększa złożoność obliczeniową. Kluczowe jest znalezienie odpowiedniego balansu, między jakością obrazu a wydajnością modelu. Styl życia i stan zdrowia również mają duży wpływ na wygląd twarzy. Osoby prowadzące zdrowy tryb życia mogą wyglądać młodziej niż osoby w tym samym wieku biologicznym, co może prowadzić do błędnych estymacji. Modele muszą uwzględniać te różnice, co wymaga dostępu do zróżnicowanych zbiorów danych. Genetyka odgrywa istotną rolę w procesie starzenia. Różnice genetyczne między osobami mogą powodować znaczne różnice w wyglądzie, co utrudnia dokładne przewidywanie wieku. Modele muszą być trenowane na danych obejmujących różne grupy etniczne, aby móc dokładnie estymować wiek niezależnie od tych różnic. Na dokładność estymacji wpływają również modyfikacje twarzy, takie jak chirurgia plastyczna, makijaż czy inne zabiegi kosmetyczne. Zmiany te mogą zaburzać naturalne cechy twarzy, co utrudnia modelom poprawne oszacowanie wieku. Modele muszą być trenowane w sposób uwzględniający takie modyfikacje lub stosować techniki przetwarzania obrazu, które potrafią "odczytać" prawdziwe cechy twarzy.

1.2 Uzyskiwanie twarzy ze zdjęcia

Rozpoznawanie twarzy jest naturalnym zadaniem dla ludzi, a eksperymenty pokazały, że nawet jednodniowe niemowlęta potrafią odróżniać znane im twarze. Wydaje się, więc, że powinno być to proste również dla komputerów. Jednak w rzeczywistości, nasza wiedza na temat tego, jak ludzie rozpoznają twarze, jest wciąż ograniczona. Czy w procesie tym ważniejsze są cechy wewnętrzne (oczy, nos, usta) czy zewnętrzne (kształt głowy, linia włosów)? Jak analizujemy obraz i jak mózg go koduje? David Hubel i Torsten Wiesel wykazali, że w naszym mózgu znajdują się wyspecjalizowane komórki nerwowe reagujące na określone lokalne cechy sceny, takie jak linie, krawędzie, kąty czy ruch. Ponieważ nie

postrzegamy świata, jako zbioru rozproszonych elementów, nasza kora wzrokowa musi jakoś łączyć różne źródła informacji w spójne wzorce. Automatyczne rozpoznawanie twarzy polega na wyodrębnieniu tych znaczących cech z obrazu, przekształceniu ich w użyteczną reprezentację i dokonaniu na niej odpowiedniej klasyfikacji (Wagner, 2012, s. 2).

Rysunek 1 Schemat wykrywania twarzy



Źródło: Opracowanie własne

Istnieje wiele różnych metod rozpoznawania twarzy, których można użyć do wyekstrahowania samej twarzy ze zdjęcia. Problem jest o tyle istotny, iż do metody estymacji wieku na podstawie zdjęcia potrzebne będą zdjęcia twarzy, a jak wiadomo nie każde zdjęcie jest robione w tej samej pozycji oraz odległości. Co za tym idzie pozyskanie dokładnego zdjęcia twarzy będzie kluczowe na etapie wczesnego przygotowywania danych. Z podstawowych metod używanych do wyodrębnienia twarzy ze zdjęcia możemy zaliczyć (Yang i in., 2002, s. 2-3):

- Metody oparte na wiedzy (Knowledge-based methods)

Metody te są w głównej mierze oparte na regułach, które zbierają naszą wiedzę na temat twarzy i starają się ją przełożyć na zestaw zasad. Przykładowo możemy sobie wyobrazić parę prostych reguł dotyczących twarzy, takich jak symetryczne rozmieszczenie oczu czy też ciemniejsze obszary wokół oczu w porównaniu z policzkami. Dużą zaletą tej metody jest jej

stosunkowa prostota, oraz łatwość w implementowaniu. Jednakże dużym problemem jest trudność w stworzeniu odpowiedniego zestawu reguł. Jeżeli reguły te będą zbyt ogólne to może to doprowadzić do sytuacji gdzie będzie dużo fałszywie pozytywnych wyników. Z drugiej strony nadmierne uszczegółowienie tych reguł prowadzić będzie do wielu fałszywie negatywnych wyników. Podejście to jest, więc trochę ograniczone, a dodatkowo nie jest w stanie znaleźć wielu twarzy na złożonym obrazie (Marques, 2010, s. 12-13).

- Metody inwariantne cech (Feature-invariant methods)

Metody te są nieco podobne do poprzednich. Starają się one znaleźć cechy twarzy, które będą stabilnie i niezależne rożnego rodzaju warunków takich jak kąt czy też pozycja twarzy. Główne założenie tych metod polega na tym, że istnieją pewne zestawy cech twarzy, które są niezależne od zewnętrznych czynników. Do takich cech możemy zaliczyć między innymi: geometrię twarzy, teksturę skóry czy też specyficzne wzorce kolorów. Jedną z najpopularniejszych metod jest Histogram of Oriented Gradients (HOG), która analizuje lokalne gradienty obrazu, tworząc histogramy, które są odporne na zmiany oświetlenia i cieni. Inna powszechnie stosowana technika to Skalowalne Niezależne od Skali Transformacje Ceł (SIFT), która identyfikuje lokalne kluczowe punkty obrazu i opisuje je za pomocą wektorów cech, inwariantnych wobec rotacji i skalowania. Local Binary Patterns (LBP) to kolejna metoda, która analizuje teksturę obrazu, przekształcając ją w binarny wzór, co pozwala na efektywną klasyfikację twarzy mimo różnic w oświetleniu.

- Metody dopasowania szablonów (Template matching methods)

Metody dopasowywania szablonów próbują zdefiniować twarz, jako funkcję. W tym celu starają się znaleźć standardowy szablon dla wszystkich twarzy. Różne cechy twarzy mogą być definiowane niezależnie, co oznacza, że możemy na przykład podzielić twarz na poszczególne atrybuty takie jak oczy, kontur twarzy nos i usta. W tej metodzie można również zastosować model twarzy opartej na krawędziach (Marques, 2010, s. 13). Proces detekcji twarzy oparty na tej metodzie jest w miarę prosty. Algorytm polega na tym, że przesuwa szablon po większym obrazie, porównując każdą sekcję z szablonem za pomocą miary podobieństwa, takiej jak korelacja czy suma bezwzględnych różnic. A by określić czy dopasowanie zostało znalezione stosuje się próg na podstawie, którego decyduje się czy wartość miary prawdopodobieństwa przekroczyło ustaloną wartość graniczną. Główne zalety metody to jej prostota i łatwość implementacji, a także skuteczność, gdy szablon i docelowy

obraz są podobne pod względem skali i orientacji. Jednakże, metoda ta jest wrażliwa na zmiany skali, obrotu oraz warunków oświetleniowych i jest dość wymagająca obliczeniowo, szczególnie w przypadku dużych obrazów lub wielu szablonów (Tan Yeh Ping i in., 2016, s. 7).

- Metody oparte na wyglądzie (Appearance-based methods)

Metody oparte na wyglądzie zostały szczególnie upodobane przez badaczy z różnych obszarów. Te sposoby w głównej mierze oparte są na technikach analizy statystycznej oraz uczenia maszynowego. Metody te możemy podzielić na 2 podejścia. Pierwszym z nich jest podejście holistyczne, które w rozpoznawaniu twarzy polega na analizie całego obrazu, jako jednej jednostki nie skupiając się na specyficznych cechach twarzy. Do jednej z najbardziej popularnych metod rozpoznawania twarzy należy zaliczyć „Principial component analysis" w skrócie (PCA). Metoda ta polega użyciu analizy głównych składowych do redukcji wymiarów danych wejściowych przekształcając je do przestrzeni o niższej wymiarowości zachowując jak najwięcej informacji o oryginalnych danych. Z drugiej strony podejście hybrydowe przy rozpoznawaniu twarzy łączy ze sobą metody holistyczne z analizą lokalnych cech twarzy. Polega to na tym, że zamiast traktować obraz, jako jedną całość, metoda hybrydowa dzieli obraz na mniejsze segmenty takie jak oczy nos i usta, a następnie analizuje te segmenty oddzielnie, by na końcu połączyć wyniki z analizy globalnej oraz lokalnej. Do tego rodzaju podejścia możemy zaliczyć „contourlet based PCA". Metoda wykorzystuje transformację Contourlet w połączeniu z PCA do ekstrakcji bardziej dyskryminacyjnych cech, co prowadzi do wyższej dokładności rozpoznawania twarzy. Podejście hybrydowe wiąże się jednak z zwiększoną złożonością obliczeniową (Parisa Beham i Mohamed Mansoor Roomi 2012, s. 16-17).

Przechodząc przez różne metody uzyskiwania twarzy ze zdjęcia można dojść do wniosku, że część z wyżej wymienionych algorytmów jest lepsza od innych jednakże wiąże się z wyższym zapotrzebowaniem na moc obliczeniową.

Z uwagi na skuteczność w różnych warunkach oraz możliwość radzenia sobie ze zmiennością w pozycjach, kątach i warunkach oświetleniowych, metody inwariantne cech oraz metody oparte na wyglądzie są najlepsze do wyodrębniania twarzy ze zdjęcia. Metody inwariantne cech są idealne tam, gdzie istnieje duża zmienność w kącie i pozycji twarzy, ale mogą mieć ograniczenia w ekstremalnych warunkach oświetleniowych. Metody oparte na

wyglądzie oferują wysoką dokładność dzięki uczeniu na dużych zbiorach danych i są szczególnie skuteczne w rozpoznawaniu twarzy w różnych warunkach, jednak wymagają dużych zasobów obliczeniowych i odpowiednich danych treningowych.

W Internecie istnieje wiele dostępnych bibliotek i kodów źródłowych, które implementują różnego rodzaju algorytmy rozpoznawania twarzy ze zdjęcia. Poniżej zostały przedstawione główne biblioteki implementujące wcześniej wymienione algorytmy w języku Python.

- Open CV - Jest to jedna z najpopularniejszych bibliotek do przetwarzania obrazów i detekcji twarzy. Umożliwia ona szbykie i łatwe wykrywanie twarzy za pomocą wbudowanych algorytmów. Zawiera ona takie metody rozpoznawania twarzy jak Haar Cascades oraz Local Binary Patterns (Dokumentacja OpenCV).
- Dlib - Dlib jest bardzo wszechstronną biblioteką Pythona, która znana jest z dokładności i wydajności w zakresie detekcji twarzy. Biblioteka ta w głównej mierze wykorzystuje zaawansowane algorytmy uczenia maszynowego. Z algorytmów potrafiących rozpoznawać twarze, które są zaimplementowane w tej bibliotece należy wymienić algorytm HOG, czyli „Histogram of Oriented Gradients", który identyfikuje obszary twarzy na podstawie gradientów i histogramów, a także algorytm oparty na konwolucyjnych sieciach neuronowych (Rosebrock, 2021).
- Face recognition - Biblioteka ta jest zbudowana na bazie Dlib i jest uważana za najprostszą w obsłudze bibliotekę do wykrywania twarzy ze zdjęcia. Pakiet ten korzysta z tych samych algorytmów, co Dlib, natomiast jest o wiele prostszy w użyciu oraz implementacji (Dokumentacja Face Recognition).
- MTCNN - Pakiet ten jest mniej popularny od reszty wcześniej wymienionych pakietów, zawiera on implementację wielozadaniowych kaskadowych konwolucyjnych sieci neuronowych, wytrenowanych na potrzebę rozpoznawania twarzy (Dokumentacja MTCNN).

1.3 Przegląd dostępnych danych

Wiele tematów związanych z uczeniem maszynowym czy też z zwykłą analizą statystyczną napotyka na ten sam problem, a mianowicie dostęp oraz jakość danych. Współcześnie problem jest prostszy do zaadresowanie głównie dzięki rozwojowi sztucznej inteligencji, nie mniej często ciężko jest zadbać o poprawne i jakościowe dane. W przypadku problemu estymacji wieku ze zdjęcia jest podobnie. Ze względu na spore różnice w indywidualnym wyglądzie twarzy oraz różnic pomiędzy ludźmi, wydajność naszego modelu w głównej mierze będzie zależna od danych treningowych. W związku z tym można dojść do wniosku, że największym wyzywaniem w tworzeniu takiego modelu będzie odpowiednia baza danych, na której sam model będzie mógł uczyć się w taki sposób, aby maksymalizować późniejsze wyniki na danych testowych (Zhang i Bao, 2022, s. 7). Przeglądając Internet można natrafić na wiele zbiorów danych zawierających zarówno zdjęcia jak i wiek osoby znajdującej się na zdjęciu. Przykładem takich danych mogą być te zbiory danych przedstawionych w tabeli numer 1.

Tabela 1 Podstawowe dane dotyczące dostępnych baz danych

Źródło danych	Średnia wieku	Przedział wiekowy	Ilość dostępnych zdjęć
IMDB-Wiki Dataset	36 lat	0-100 lat	523,051
UTKFace Dataset	34 lat	0-116 lat	24,102
Adience Dataset	25-34 lat	0-60+ lat	26,580
MORPH Dataset	39 lat	16-77 lat	55,134
CACD (Cross-Age Celebrity Dataset)	33 lat	14-62 lat	163,446
APPA-REAL Dataset	32 lat	0-95 lat	7,591
MegaAge	23 lat	0–70 lat	41,941
Facial Recognition Technology (FERET)	18-30 lat	1–66	14,126
FGNET	16 lat	0–69	1002
YGA	(-) lat	0–93	8000
Images of Group (IoG)	(-) lat	0–66	5080

Źródło: Opracowanie własne na podstawie danych z Internetu

Każdy z wymienionych zestawów danych ma swoje unikalne cechy i zastosowania, które mogą być wykorzystane w badaniach nad oceną wieku. Wybór odpowiedniego zestawu danych zależy od specyficznych potrzeb badawczych, takich jak różnorodność demograficzna, ilość danych, czy specyfika zadania.

Można oczywiście połączyć ze sobą te zbiory danych, jednak taka operacja wymaga zdecydowanie większej mocy obliczeniowej i może wymagać sporych zasobów, aby przetestować model na tak dużej ilości danych. Jednakże takie podejście może znacząco polepszyć, jakość modelu. Wykorzystanie głębokich sieci neuronowych w połaczeniu z podejściem „corss-dataset" a więc wykorzystywanie wielu zbiorów danych może znacząco poprawić dokładność estymacji wieku na podstawie zdjęcia (Kuang i in., 2015, s. 100).

1.4 Przegląd literatury

Automatyczne rozpoznawanie wieku za pomocą sieci neuronowych staje się coraz bardziej popularnym tematem badawczym. Postęp technologiczny w zakresie budowania takich modeli oraz łatwy dostęp do różnorodnych „frameworków" znacząco przyczyniły się do wzrostu zainteresowania tym zagadnieniem. W ostatnich latach pojawiło się wiele badań i konkursów, które zachęcają do rozwijania efektywnych metod estymacji wieku z obrazów twarzy. Dzięki temu badacze i inżynierowie mają możliwość eksperymentowania z nowymi technikami, co prowadzi do ciągłego doskonalenia i zwiększania dokładności modeli predykcyjnych.

Przy tworzeniu modeli szacujących wiek na podstawie obrazu należy uwzględnić wiele aspektów. Do kluczowych etapów i wyzwań związanych z budową takiego modelu możemy między innymi zaliczyć:

- Pozyskiwanie danych: Ważne jest odpowiednie dobranie danych treningowych i testowych, które będą reprezentatywne dla różnych grup wiekowych i warunków.
- Wstępne przetwarzanie: Obejmuje to ekstrakcję twarzy, jej wyrównanie oraz normalizację, aby przygotować zdjęcia do dalszej analizy.
- Ekstrakcja cech: Proces ten polega na wyodrębnianiu cech związanych z procesem starzenia, takich jak zmarszczki czy struktura kości twarzy.

- Trenowanie modelu: Używa się różnych klasyfikatorów i technik uczenia maszynowego do trenowania modelu, który będzie przewidywać wiek na podstawie zdjęć.

Należy podkreślić również istotę poprawnych danych, oraz implementacji różnych technik, aby zwiększyć dokładność modeli szacujących wiek. Ważne jest także radzenie sobie z wyzwaniami, takimi jak różnorodność danych, unikalne wzorce starzenia oraz jakość zdjęć, aby uzyskać jak najlepsze wyniki (Elkarazle, 2022, s. 99-101).

Dodatkowo warto zwrócić uwagę na wpływ wewnętrznych czynników przy estymacji taki jak np. tożsamość, płeć, rasa czy też stan zdrowia. Kolejnym ważnym elementem są czynniki zewnętrzne takie jak np. pozowanie twarzy, okulary czy makijaż. Co ważne należy przede wszystkim podkreślić znaczenie kompleksowego podejścia do estymacji wieku, uwzględniającego różne techniki ekstrakcji cech, różnego rodzaju algorytmy oraz bazy danych (Al-Shannaq i Elrefaei, 2019, s. 18).

Do przeprowadzenia oszacowania wieku na podstawie zdjęć można podejść na wiele sposobów. Jednym z nich jest posegmentowanie wieku na różne kategorie, tak, aby model finalnie nie przewidywał dokładnego wieku, a umieszczał poszczególne zdjęcia w odpowiedniej kategorii, np. wiek w przedziale (10,15). Eksperymenty wykazały, że przy zastosowaniu konwolucyjnych sieci neuronowych uzyskano dokładność wynoszącą około 0,52. Dodatkowo, model ten został porównany z innym stworzonym na innym zbiorze danych. Wyniki w tym porównaniu wypadły jednak słabo, uzyskując średnio o 20 p.p. gorszą dokładność niż model porównawczy. Rezultaty jasno wskazują, że część grup wiekowych była lepiej przewidywana niż inne, co prowadzi do wniosku, że przedziały wiekowe powinny być lepiej dopasowane w zbiorze testowym oraz treningowym (Kjærrani i in., 2022, s. 10). Ponadto, wnioski podkreślają, że wynik można byłoby polepszyć przy zastosowaniu "transfer learningu", czyli techniki polegającej na wykorzystaniu wiedzy zdobytej przez inny model (Torrey i Shavlik, 2019, s. 24).

Podobne wnioski zostały zaprezentowane w badaniach przeprowadzonych przez Antonio Greco, Alessię Saggese, Mario Vento oraz Vincenzo Vigilante. Autorzy przeprowadzili tak zwaną destylację wiedzy, która polega na trenowaniu mniejszego, prostszego modelu (tzw. modelu studenta), bazującego na wynikach bardziej złożonego modelu (tzw. modelu nauczyciela). Metoda ta sprawia, że możliwe jest stworzenie modelu równie wydajnego i dokładnego, ale wymagającego zdecydowanie mniej zasobów

obliczeniowych. Uzyskany w ten sposób model studenta działał nawet 15 razy szybciej niż model nauczyciela. Sam w sobie nie był dokładniejszy niż model nauczyciela, ale porównywalny z innymi bardziej złożonymi metodami. Co ciekawe, w przypadku różnych zakłóceń obrazu (np. rozmycie, szumy, zmiany jasności) model studencki radził sobie lepiej niż model nauczyciela. Wyniki sugerują, że destylacja wiedzy jest bardzo skuteczną techniką estymacji wieku (Greco i in., 2022, s. 12-13).

Kolejnym sposobem na tworzenie modelu szacującego wiek na podstawie zdjęcia może być podejście wykorzystujące technikę rozkładów etykiet. W tym podejściu do każdego obrazu przypisywany jest rozkład etykiet zamiast jednej etykiety wieku. W dużym uproszczeniu, dane zdjęcie zamiast otrzymywać tylko predykcję na poziomie np. (10-15 lat), otrzymuje szereg predykcji z przypisanymi wartościami prawdopodobieństwa. Dzięki temu jedno zdjęcie twarzy przyczynia się do nauki modelu również dla kilku sąsiednich grup wiekowych. Eksperymenty wykazały, że takie podejście oparte na rozkładach etykiet może być bardziej efektywne niż tradycyjne metody. Dodatkowo, metoda ta może być znacznie lepsza, jeżeli zasoby danych treningowych są ograniczone lub rozproszone w szerokim zakresie wiekowym (Geng i in., 2013, s. 5-6).

W 2021 został przeprowadzony konkurs „Guess The Age, 2021" który miał na celu opracowanie metod estymacji wieku z obrazów twarzy, używając nowoczesnych rozwiązań takich jak głębokie sieci neuronowe. Organizatorzy konkursu dostarczyli uczestnikom ogromny zbiór danych zawierających 575 tysięcy zdjęć z etykietami wieku. Jest to jeden z największych aktualnie dostępnych takich zbiorów danych. Zasady konkursu były proste, poszczególne zespoły miały dostarczyć model, który maksymalizuje dokładność na ustalonym wcześniej zbiorze testowym, przy założeniu, że zespoły korzystają tylko i wyłącznie z zestawu treningowego. Wydajność metod oceniano na podstawie indeksu AAR (Age Accuracy and Regularity), który łączy w sobie średni błąd absolutny a także odchylenie standardowe błędu. Wzór na liczenie tego indeksu znajduje się poniżej.

$$\mathrm{AAR} = \max(0; 7 - \mathrm{MAE}) + \max(0; 3 - \sigma)$$

$$\mathrm{MAE} = \frac{1}{K}\sum_{i=1}^{K}|p_i - r_i|$$

$$\sigma = \frac{1}{K}\sqrt{\frac{1}{8}\sum_{j=1}^{8}\left(MAE_j - MAE\right)^2}$$

Do konkursu początkowo przystąpiło 20 drużyn, z których tylko 7 przesłało finalną wersję modelu. Zadaniem każdego zespołu było uzyskanie jak najwyższego wskaźnika ARR. Wyniki poszczególnych drużyn zostały przedstawione w tabeli poniżej.

Tabela 1 Wyniki konkursu „Guess the Age"

Pozycja w rankingu	Zespół	Baza danych	Model	ARR
1	BTWG	MS-Celeb-1M	EfficientNetV2-M	7,94
2	Pacific of Artificial Vision	VGGFace-2	ResNet-50	7,55
3	CIVA Lab	ImageNet	ResNeXt	6,97
4	Levi	Nieznana	DeepFace	5,64
5	VisionH4ck3rz	ImageNet	EcientNet-B0	5,41
6	GoF	ImageNet	ResNet-50	3,8
7	GvisUleTeam	Nieznana	ResNet-50	3,69

Źródło: Greco, 2021, s. 8

- BTWG – zespół ten wykorzystał sieć EfficientNetV2-M, która została wstępnie wytrenowana na zbiorze danych MS-Celeb-1M, a następnie dostrojona do szacowania wieku na zbiorze danych MIVIA. Procedura uczenia została podzielona na dwa etapy. Pierwszy etap zakładał uczenie reprezentacji a następnie klasyfikację. Na etapie przygotowywania danych zespół przeprowadził augmentację danych za pomocą RandAugment i reprezentował etykiety wieku, jako rozkłady normalne. W kolejnym etapie model został wytrenowany na cały zbiorze treningowym z zastosowaniem niestandardowej funkcji straty łączącej dywergencję KL i stratę L1 w celu regularyzacji. W ostatnim etapie, czyli klasyfikacji, zespół dostroił w pełni połączone warstwy oraz warstwy wyjściowe, przy użyciu zrównoważonej wersji zestawu treningowego oraz zmodyfikowanej funkcji straty MSE.
- Pacific of Artificial Vision – drużyna wykorzystała model ResNet-50, który został wstępnie wytrenowany na zbiorze VGGFace2 do rozpoznawania twarzy, a następnie dostrojony do oszacowania wieku na rozszerzonym zbiorze treningowym.

Rozszerzenie zakładało wzbogacenie zbioru o dodatkowe zdjęcia odzwierciedlające proces starzenia. Rozszerzenie obejmowało mniej reprezentatywne grupy wiekowe. Zastosowano niestandardową funkcję straty, będącą kombinacją średniej wariancji i krzyżowej entropii, w celu dalszej regulacji wyników w różnych grupach wiekowych.

- CIVA Lab – w tym podejściu został zastosowany model ResNeXt, który wstępnie został przetrenowany na zbiorze danych ImageNet. W kolejnym kroku został on dostrojony do estymacji wieku na całym zestawie treningowym, który został wcześniej znormalizowany, wyrównany oraz uzupełniony losowym odbiciem poziomym każdego zdjęcia. Zespół postawił na niestandardową funkcję strat, która była inspirowana indeksem AAR. Uzyskane osadzenie 2048 cech zostało podane, jako dane wejściowe do dwuwarstwowego lasu losowego (TLRF) z 100 drzewami w każdej warstwie, który działał, jako regresor, zwracając przewidywany wiek.
- Levi – wytrenował model DeepFace od podstaw do estymacji wieku (DeepAge), usuwając warstwę przedostatnią w celu zmniejszenia złożoności obliczeniowej. Aby zredukować nierównowagę zbioru treningowego, zastosowali autoenkoder do generowania dodatkowych próbek dla mniej reprezentowanych grup wiekowych, uzyskując 1000 próbek dla tych grup i 2000 próbek dla pozostałych. Dodatkowo, zastosowano techniki augmentacji, takie jak kadrowanie, zmiana jasności, kontrastu, rozmycie, konwersja do skali szarości, przesunięcie, skalowanie i obrót, aby rozszerzyć zbiór danych.
- VisionH4ck3rz – zespół zastosował rozszerzenie danych o losowe odbicie zdjęć, powiększenie, obrót oraz zmiana jasności. Rozszerzenie takie zostało zaaplikowane tylko do mniej reprezentatywnych grup wiekowych, aby uzyskać, co najmniej 8000 próbek dla każdej grupy. W kolejnym kroku przy użyciu tego zestaw danych oraz funkcji straty MAE został wstępnie przetrenowany model EfficientNet-B0. Ostatecznie został dostrojony poprzez zastąpienie oryginalnej warstwy wyjściowej globalnym średnim poolingiem, normalizacją wsadową, warstwą dropout oraz pojedynczym neuronem wyjściowym z liniową aktywacją.
- GoF - Zespół zmodyfikował model ResNet-50, wstępnie wytrenowany na zbiorze ImageNet, dodając dwie dodatkowe w pełni połączone warstwy i dostroił go na zestawie treningowym przy użyciu funkcji straty MSE. Nie przeprowadzono rozszerzenia danych, lecz obrazy zostały wstępnie przetworzone poprzez zastosowanie kanonicznego wyrównania twarzy.

- GvisUleTeam - Zespół wyszkolił model SSR-Net od podstaw, korzystając z funkcji straty MAE. Aby zmniejszyć nierównowagę w zestawie treningowym, użyli StyleGANv2 i HRFAE, aby uzyskać, co najmniej 4000 próbek dla nieletnich i starszych grup wiekowych. Dodatkowo, aby zwiększyć reprezentatywność danych, zastosowali różne techniki rozszerzania danych, takie jak losowe maskowanie, obracanie i powiększanie.

Konkurs pokazał, że sukces w estymacji wieku zależy od połączenia odpowiedniej strategii augmentacji danych, zastosowania niestandardowych funkcji strat oraz wykorzystania zaawansowanych modeli neuronowych, najlepiej wstępnie wytrenowanych. Najlepsze wyniki uzyskano, kiedy wszystkie te elementy były harmonijnie połączone, co pozwoliło na stworzenie modeli o wysokiej dokładności i stabilności, nawet w przypadku trudnych do klasyfikacji grup wiekowych (Greco, 2021, s. 4-9).

1.5 Kwestie etyczne i prywatność związane z wykorzystaniem danych

Szacowanie wieku na podstawie twarzy jest potężnym narzędziem, które może znacząco wspierać działania na rzecz bezpieczeństwa publicznego i egzekwowania prawa. Zdolność tej technologii do identyfikacji i weryfikacji osób jest szczególnie cenna w sytuacjach, gdy brak jest odpowiednich dokumentów tożsamości (Rahman i in., 2023, s. 1-2). Rozpoznawanie twarzy i ocena wieku za pomocą technologii komputerowych budzą wiele kontrowersji związanych z kwestiami etycznymi i prywatnością. Jednym z kluczowych problemów związanych z technologią rozpoznawania twarzy jest gromadzenie danych bez wiedzy i zgody użytkowników. Często wykorzystuje się obrazy pobrane z Internetu, bez zgody osób, które na nich widnieją. Choć w niektórych jurysdykcjach jest to legalne, takie praktyki coraz częściej uznawane są za nieetyczne. Niezbędne jest wprowadzenie regulacji, które będą wymagały transparentności oraz zgody na zbieranie i przetwarzanie tego typu danych (Girasa, 2020, s. 7-8). Kolejnym problemem etycznym, jaki powstaje w tego typu systemach jest sytuacja, w której algorytmy rozpoznawania twarzy mogą być obciążone uprzedzeniami prowadząc do dyskryminacji różnych grup demograficznych. Badania wykazały, że algorytmy te często lepiej rozpoznają twarze osób białych w porównaniu do osób o ciemniejszej karnacji (Turner Lee i Chin-Rothmann, 2022).

Rozdział 2 Opis metodyki oraz przegląd Modeli

2.1 Wstępne przygotowanie danych

Konwertowanie zdjęć na dane wejściowe do modeli uczenia maszynowego jest bardzo ważnym elementem w wielu podejściach związanych z budowaniem jakiegokolwiek modelu. Proces ten obejmuje szereg kroków od wczytania obrazu poprzez wstępną obróbkę a na ekstrakcji cech oraz augmentacji danych kończąc. Dane do modeli mogą być dostępne w wielu formach. Zaczynając do najprostszych uporządkowanych danych kończąc na zdjęciach filmach czy też plikach audio. Komputer nie jest w stanie bezpośrednio zinterpretować zdjęcia czy też nagrania wideo, ponieważ może on rozpoznawać tylko wartości 0 oraz 1. Zatem w celu budowania modeli należy najpierw zamienić dostępne zdjęcia na odpowiednie macierze w taki sposób, aby dostępne algorytmy mogły dać jakąkolwiek wartość dodatnią. Pierwszym krokiem w przygotowywaniu danych będzie wczytanie obrazu. Każdy cyfrowy obraz składa się z pikseli, które są najmniejszymi jednostkami obrazu. W przypadku obrazu cyfrowego każdy piksel jest reprezentowany przez trzy wartości odpowiadające intensywności kolorów w przestrzeni RGB, czyli odpowiedni czerwony zielony oraz niebieski. W obrazach w skali szarości, każdy piksel jest reprezentowany przez jedną wartość, a dokładniej przez intensywność szarości. W przypadku wczytywania obrazów, jako danych do modeli każde zdjęcie jest reprezentowane przez macierz o wymiarach W x H, gdzie W oznacza szerokość natomiast H wysokość wyrażoną w liczbach odpowiadających liczbom pikseli. Jeżeli jest to zdjęcie kolorowe to dodatkowy 3 wymiar jest odpowiedzialny za 3 różne kanały RGB. Przekształcenie zdjęcia w odcienie szarości można przeprowadzić na wiele sposobów. Proces ten polega na zastosowaniu liniowej kombinacji wagowej wartości kanałów R (czerwony), G (zielony) i B (niebieski) dla każdego piksela obrazu. Wzór na konwersję obrazu z formatu RGB do skali szarości pochodzący z biblioteki TensorFlow jest następujący (Dokumentacja TensorFlow):

$$Y = 0{,}2989 * R + 0{,}5870 * G + 0{,}1140 * B$$

gdzie:

- R to wartość kanału czerwonego (Red),
- G to wartość kanału zielonego (Green),
- B to wartość kanału niebieskiego (Blue),
- Y to wartość jasności (luminancji) w odcieniach szarości.

Rysunek 2 Reprezentacja zdjęcia w formacie macierzowej

Źródło: Opracowanie własne na podstawie zdjęcia z Internetu

Zamiana obrazu na odcień szarości zdecydowanie upraszcza algorytm a dzięki temu wymaga mniej mocy obliczeniowej ze względu na mniej wymiarów w macierzy. Jednak jak pokazały badania w zadaniach, gdzie kolor odgrywa kluczową rolę, kolorowe obrazy mogą prowadzić do lepszych wyników. W przypadku ograniczeń sprzętowych lub gdy kolor nie jest istotny, szare obrazy mogą być bardziej efektywne (Kanan i Cotterell, 2012, s. 3-4). W przypadku problemu oceny wieku ze zdjęcia kolory nie będą dodawać żadnej dodatkowej wartości informacyjnej a będą jedynie sprawiać, że finalny algorytm będzie bardziej skomplikowany.

Kolejnym ważnym krokiem w procesie przygotowywanych jest zmiana rozmiaru obrazów, które będą znajdować się w zbiorze treningowym. Zmiana rozmiaru obrazów jest kluczowym krokiem w przetwarzaniu danych wejściowych dla modeli uczenia maszynowego, zwłaszcza w zadaniach takich jak estymacja wieku ze zdjęć. Proces ten polega na skalowaniu obrazów do jednolitego rozmiaru, co pozwala na spójne i efektywne przetwarzanie danych przez model. Analizy wpływu zmiany rozmiaru obrazów na czas treningu i wydajność modeli detekcji obiektów opartych na głębokim uczeniu wskazują, że w skalowaniu obrazów należy zachować kompromis pomiędzy czasem a dokładnością (Saponara i Elhanashi, 2022, s. 6-7).

W przypadku estymacji wieku ze zdjęcia można stwierdzić, że znaczne zmniejszanie rozmiaru zdjęć może prowadzić do utraty dokładności w modelu, jednak badania pokazują, że straty te nie są duże.

Jak pokazał wcześniejszy przegląd literatury odnośnie tworzenia modeli estymacji wieku ze zdjęcia bardzo ważnym elementem w procesie przygotowywania danych jest proces augmentacji danych. Polega on na sztucznym powiększeniu zbioru treningowego w taki sposób, aby sama konstrukcja oraz zawartość danych nie zaburzała procesu uczenia modelu. Jest wiele technik augmentacji danych, z którymi można się spotkać w przypadku modeli wykorzystujących zdjęcia. Do najprostszych technik możemy zaliczyć te, które oparte są na transformacji geometrycznej. Transformacja polega na odwracaniu obrazu w pionie lub poziomie, aby uzyskać lustrzane odbicie, rotację obrazu o określony kąt, przesuwanie obrazu a także przycinanie oraz skalowanie. Transformacja będzie oparta na wcześniej istniejących zdjęciach a nowo powstałe zdjęcia zwiększą próbę treningową. Kolejnym sposobem na uzyskanie dodatkowych zdjęć będzie transformacja w przestrzeni barw. W tym przypadku możemy wyszczególnić takie metody jak manipulacje w przestrzeni kolorów RGB, zmiany jasności, kontrastu i nasycenia oraz przekształcenia histogramów kolorów. Do trochę bardziej zaawansowanych metod możemy zaliczyć filtry jądra tak zwane „Kernel filters". Technika ta polega na przetwarzaniu obrazu poprzez modyfikację wartości pikselu na podstawie wartości sąsiednich pikseli. Są stosowane w wielu operacjach takich jak wygładzanie, wyostrzenie czy też detekcja krawędzi. Do głównych metod możemy zaliczyć „Gaussian Blur" oraz „Sharpening". Gaussian Blur to technika wygładzania obrazu polegająca na zastosowaniu filtru Gaussa, który rozmywa obraz, eliminując szum i drobne szczegóły. Rozmycie jest osiągane przez splot obrazu z funkcją Gaussa, co zapewnia miękkie i płynne przejścia między pikselami. Z kolei sharpening polega na zwiększaniu kontrastu między sąsiednimi pikselami, co uwydatnia szczegóły i krawędzie. Filtry wyostrzające, takie jak Laplacian lub filtry wysokoprzepustowe, są używane do tego celu. Kolejnym sposobem, w jaki można dokonać augmentacji danych będzie mieszanie obrazów. Sposób ten polega na łączeniu dwóch obrazów poprzez uśrednianie wartości pikseli. Każdy piksel w nowym obrazie jest średnią wartością odpowiadających mu pikseli w dwóch oryginalnych obrazach. Ta technika jest używana do tworzenia bardziej zróżnicowanych zestawów danych, co pomaga modelom w nauce generalizacji i radzeniu sobie z różnorodnymi przypadkami (Shorten i Khoshgoftaar, 2019, s. 7-23).

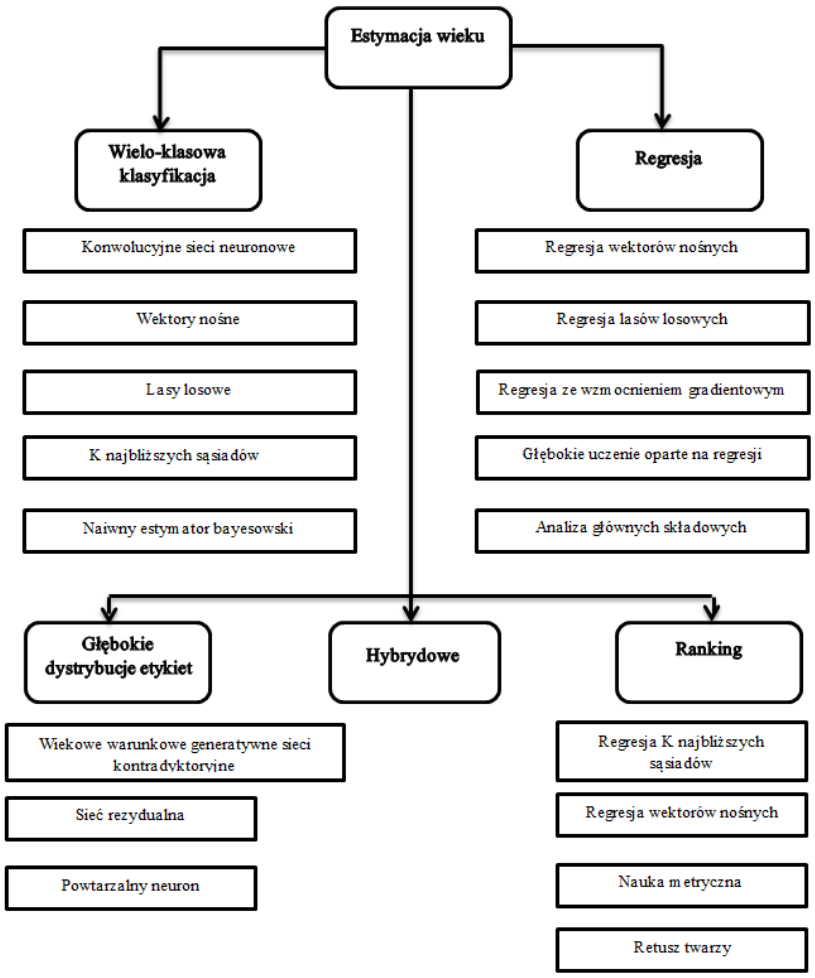
2.2 Przegląd dostępnych modeli oraz funkcji celu

Do problemu estymacji wieku ze zdjęcia można podejść na wiele sposób. W najnowszej literaturze najczęściej pojawiają się wszelkiego rodzaju techniki związane z głębokimi sieciami neuronowymi, które są trenowane na sporych zbiorach danych. Jednakże nie są to jedyne metody, które można zaaplikować do tego problemu. Problem oceny wieku ze zdjęcia można sprowadzić nie tylko do zastosowanego algorytmu, ale również do ogólnej metody, jaka zostanie zastosowana w procesie modelowania. Te metody możemy podzielić na (Ghrban i Abbadi, 2023, s. 9):

- Wielo-klasowa klasyfikacja – Wielo-klasowa klasyfikacja traktuje estymację wieku, jako problem klasyfikacyjny, w którym każda możliwa wartość wieku jest traktowana, jako osobna klasa. Chociaż jest prostsze do zaimplementowania, ma tendencję do problemów z przeuczaniem (overfitting) i niestabilnym treningiem, szczególnie przy większej liczbie klas.
- Regresja – Regresja jest popularną metodą wykorzystywaną w estymacji wieku, która modeluje wiek, jako zmienną ciągłą. Modele regresji minimalizują błąd, taki jak MAE (Mean Absolute Error), aby zbliżyć przewidywane wartości wieku do wartości rzeczywistych. Choć regresja może być dokładniejsza niż klasyfikacja, jest wrażliwa na wartości odstające, co może destabilizować proces treningowy, ponieważ traktuje wiek liniowo.
- Głębokie dystrybucje etykiet – Metoda ta stara się rozwiązać problemy z nierównomiernym rozkładem etykiet wiekowych, co często pojawia się w bazach danych. DLDL przyjmuje podejście, które przekształca rozkład etykiet wiekowych na bardziej równomierny i optymalny dla treningu sieci neuronowych. Jest to skuteczne rozwiązanie problemów z małą liczbą danych, jednakże może być podatne na niestabilności w procesie treningu.
- Hybrydowe – Podejścia hybrydowe łączą różne techniki, takie jak klasyfikacja i regresja, aby poprawić dokładność i odporność modeli estymacji wieku. Chociaż łączenie różnych modeli może przynieść lepsze wyniki, zwiększa również koszty obliczeniowe i jest trudniejsze do wdrożenia na urządzeniach o ograniczonych zasobach.
- Ranking - Metoda rankingowa używa osi wiekowej do przewidywania wieku na podstawie klasyfikacji binarnej, gdzie każda klasa reprezentuje przedział wiekowy.

Dzięki temu podejściu model może lepiej zrozumieć relacje między różnymi grupami wiekowymi. Jednakże, prowadzi to czasami do niespójności w treningu i ocenie, a także może być suboptymalne w porównaniu do innych metod.

Rysunek 3 Ogólne metody estymacji wieku



Źródło: Ghrban i Abbadi, 2023 s. 9

Powyższy rysunek przedstawia ogólne metody, które można zaaplikować do problemu oceny wieku ze zdjęcia twarzy. Każda z tych metod pozwala na zastosowanie konkretnego rodzaju algorytmu, którego wytrenowane parametry posłużą do późniejszego szacowania wieku na podstawie danych testowych. Algorytmy te mogą być oparte na ręcznie opracowanych cechach lub wykorzystywać sieci neuronowe.

Modele oparte na ręcznie opracowanych cechach zazwyczaj wykorzystują filtry, takie jak Histogram Zorientowanych Gradientów (HOG) lub Lokalny Wzorzec Binarny (LBP), do wyodrębniania krawędzi i kształtów z obrazu twarzy. Następnie dodaje się wybrany algorytm

uczenia, na przykład k-najbliższych sąsiadów lub maszynę wektorów nośnych, aby nauczyć się rozpoznawać wyodrębnione cechy. W ręcznie opracowanych modelach cechy starzenia są wyodrębniane manualnie za pomocą filtrów takich jak Gabor, HOG czy Sobel. Filtry te są dostosowywane, aby uzyskać jak najwięcej cech, takich jak zmarszczki, kształt głowy, tekstury czy krawędzie, które mogą wskazywać wiek osoby. Choć modele ręcznie opracowane wymagają mniejszej mocy obliczeniowej niż modele oparte na głębokim uczeniu, ich dokładność jest zwykle niższa. W artykule omówiono cztery główne podejścia: modele antropometryczne, modele oparte na teksturze, aktywne modele wyglądu (AAM) oraz podprzestrzeń wzorców starzenia (AGES) (Elkarazle i in., 2022, s. 7-8).

- Modele antropometryczne: Wykorzystują pomiary struktury ciała, aby zrozumieć geometrię ludzkiego ciała i różnicować grupy wiekowe oraz płcie.
- Modele oparte na teksturze: Zamiast odległości między punktami na twarzy, bazują na intensywności pikseli, aby wyodrębniać cechy takie jak zmarszczki.
- Aktywne modele wyglądu (AAM): Łączą modele antropometryczne i oparte na teksturze, ale mogą tracić niektóre cechy starzenia podczas redukcji wymiarowości.
- Podprzestrzeń wzorców starzenia (AGES): Identyfikuje wzorce starzenia na podstawie zestawu zdjęć twarzy ułożonych chronologicznie.

Modele wykorzystujące sieci neuronowe są najbardziej popularnymi algorytmami, które są stosowane w literaturze podejmującej problem oceny wieku ze zdjęcia. Struktura sieci neuronowych stosowanych do estymacji wieku pozostaje niezmienna, niezależnie od liczby czy rodzaju warstw. Każdy model oparty na głębokim uczeniu posiada warstwę wejściową, której rozmiar jest dopasowany do rozmiaru przetwarzanych obrazów. Na przykład, sieć z warstwą wejściową o wymiarach 96 x 96 x 3 będzie przyjmować obrazy RGB o rozdzielczości 96 x 96 pikseli. Z kolei sieć z warstwą wejściową o rozmiarze 128 x 128 x 1 będzie akceptować jedynie obrazy w skali szarości o rozdzielczości 128 x 128 pikseli. Po warstwie wejściowej dodawane są warstwy przetwarzające, które odpowiadają za ekstrakcję i analizę cech obrazu. Te warstwy nazywane są "warstwami ukrytymi", a ich liczba jest zazwyczaj ustalana na podstawie eksperymentów mających na celu znalezienie najlepszej architektury modelu. Zwykle są to warstwy konwolucyjne (CNN), ponieważ są one szczególnie efektywne w wyodrębnianiu cech i krawędzi z obrazów (Elkarazle, 2022, s. 11). Modele głębokiego uczenia mogą być trenowane od podstaw lub bazować na modelach wstępnie wytrenowanych.

- Modele budowane od podstaw - Jednym ze sposobów wyodrębniania cech twarzy z użyciem głębokiego uczenia jest opracowanie algorytmu głębokiego uczenia od podstaw. W tym podejściu można zdefiniować sieć neuronową, która składa się z kilku warstw konwolucyjnych, warstw dropout, funkcji aktywacji, warstw pooling oraz w pełni połączonych warstw. Warstwy konwolucyjne i pooling tworzą mapy cech, a warstwy dropout są stosowane, aby zapobiegać przeuczeniu przez losowe wyłączanie wybranych neuronów. Sieć w pełni połączona otrzymuje wyodrębnione cechy i uczy się funkcji odwzorowania. Trenowanie sieci od podstaw może być jednak kosztowne pod względem obliczeniowym i czasochłonne, ponieważ wymaga ciągłego dostrajania parametrów, a sieć może rosnąć wykładniczo.
- Modele wstępnie wytrenowane - Modele wstępnie wytrenowane są bardziej efektywną czasowo i przestrzennie alternatywą dla modeli głębokiego uczenia budowanych od podstaw. Wykorzystuje się w nich sieć, która została już wytrenowana na bardziej złożonym zadaniu, a następnie dostosowuje się ją do wyodrębniania cech z obrazu. Hiperparametry modelu są zazwyczaj dostrajane i modyfikowane. Przykładami wstępnie wytrenowanych modeli, które osiągnęły najwyższą dokładność w zadaniach rozpoznawania twarzy, w tym estymacji wieku, są VGG-16, VGG-19, ResNet50 i AlexNet (Elkarazle i in., 2022, s. 7-8).

Ocena wydajności modelu jest kluczowym elementem w każdym problemie, do którego zastosujemy rozwiązanie z dziedziny uczenia maszynowego. Istnieje wiele metod mających na celu ocenienie wydajności modeli szacowania wieku a każda z nich zależy od architektury, założeń oraz przeznaczenia modelu. Ponieważ estymacja wieku ze zdjęcia może być traktowana zarówno jak problem regresji jak i klasyfikacji, do wyboru jest szeroki wachlarz funkcji kosztu, które mogą zostać zastosowane do oceny wydajności modeli. Średni błąd bezwzględny (MAE) i średni błąd kwadratowy (MSE) są dostępne dla zadań regresyjnych, podczas gdy dokładność (accuracy) i skumulowany wynik (cumulative score) są używane w problemach klasyfikacyjnych (Elkarazle i in., 2022, s. 9).

Definicja MAE, która jest głównie stosowana do oceny modeli regresyjnych, jest następująca:

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n}$$

Gdzie Y reprezentuje przewidywany wiek, X to rzeczywisty wiek, a N to liczba obrazów.

Z kolei MSE jest definiowany, jako:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - X_i)^2$$

Podobnie jak w pierwszym równaniu Y reprezentuje przewidywany wiek, a X to rzeczywisty wiek elementu i. W obu przypadkach niższe wartości MAE lub MSE oznaczają, że model działa dobrze, podczas gdy wyższe wartości wskazują na duży margines błędu, co sugeruje słabszą wydajność modelu.

Jeśli model jest trenowany na wielu klasach wieku, równania MAE i MSA mogą okazać się nieodpowiednie. W takim przypadku bardziej odpowiednia może być metryka taka jak skumulowany wynik (CS) lub też dokładność (Accuracy), którego równania przedstawiają się następująco:

- Dokładność

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Gdzie TP (True Positives) oznacza liczbę prawdziwych pozytywnych wyników, czyli przypadków poprawnie zaklasyfikowanych, jako pozytywne, natomiast TN (True Negatives liczbę poprawnie przewidzianych negatywnych. Z kolei w mianowniku znajdują się liczna wszystkich predykcji. Należy jednak pamiętać, że ta metoda oceny modelu będzie nadawać się tylko wtedy, jeżeli do oceny zostaną zastosowane tylko dwie kategorie na przykład w przypadku, gdy chcemy zdeterminować czy dana osoba jest pełnoletnia. W przypadku wielu kategorii możemy wyznaczyć wartość CS.

- Cumulative Score

$$CS = \frac{n}{N} * 100\%$$

Gdzie n to liczba poprawnie sklasyfikowanych obrazów, a N to całkowita liczba obrazów testowych. Równanie CS można przekształcić, aby obliczyć liczbę poprawnie sklasyfikowanych obrazów, gdzie błąd nie przekracza określonej wartości wyrażonej w latach. Nowe równanie CS wygląda następująco:

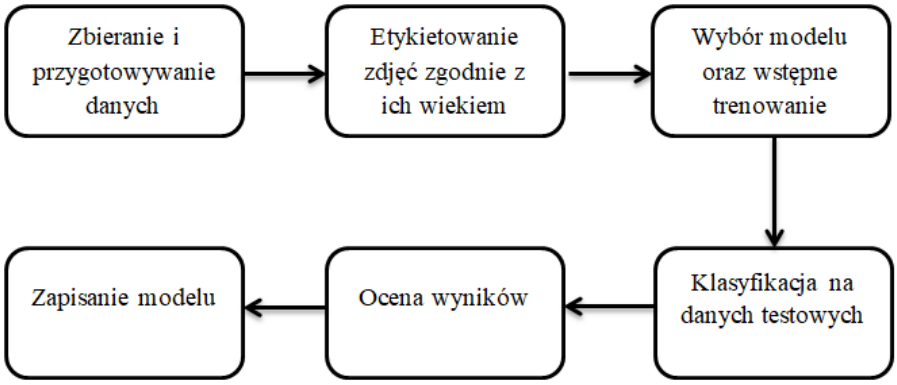
$$CS = \frac{N_{e \leq j}}{N} * 100\%$$

Gdzie N nadal oznacza całkowitą liczbę obrazów testowych, a $N_{e<=j}$ to liczba poprawnie przewidzianych obrazów, w których błąd nie przekracza j lat.

2.3 Proponowana metodyka tworzenia modelu

Przez ostatnie dwa rozdziały zostały zaprezentowane poszczególne etapy oraz ogólne metody aplikowane do tych etapów. Zgodnie z przeglądem najnowszych artykułów naukowych oraz ogólnymi obserwacjami, tworzenie modelu estymującego wiek zostanie przeprowadzone zgodnie z zaprezentowaną poniżej metodyką.

Rysunek 4 Proces modelowania i estymacji wieku ze zdjęcia



Źródło: Opracowanie własne

- Zbieranie i przygotowywanie danych

Po dokładnym przeanalizowaniu dostępnych zbiorów danych oraz wstępnej analizie, w pracy zostanie wykorzystana baza danych UTK Face. Wybór tego zbioru danych wynika z najszerszego możliwego zakresu wieku oraz ilości dobrej, jakości zdjęć. Po pobraniu danych zostaną one przygotowane do dalszego procesu modelowania. W tym celu zostaną zaznaczone twarze poszczególnych osób oraz przekształcone do odcienia szarości. Aby

zwiększyć próbę treningową oraz uniknąć potencjalnych problemów związanych z niereprezentatywnością poszczególnych grup, zostanie zastosowana augmentacja danych w taki sposób, aby uzyskać jak najlepszą próbę do treningu.

- Etykietowanie zdjęć zgodnie z ich wiekiem

Zgodnie z przyjętym założeniem etykiety zostaną przypisane zdjęciom zgodnie z wiekiem osób obecnych na zdjęciach. W tym przypadku osoby będące osobami pełnoletnimi tzn. mające więcej niż 18 lat dostaną wartość 1, natomiast w innym przypadku otrzymają wartość 0. Tak przygotowane etykiety będą podstawą do tworzenia modelu oceniającego czy dana osoba jest osobą pełnoletnią.

- Wybór modelu oraz wstępne trenowanie

Spośród wielu dostępnych metod oraz modeli, w tej pracy zostaną przetestowane dwa podejścia. Pierwsze używające wyłącznie nowo stworzonego modelu a także podejście zakładające użycie wcześniej wytrenowanego modelu w taki sposób, aby zmaksymalizować funkcję celu. Bazując na przeglądzie literatury do modelowania zostaną zastosowane konwolucyjne sieci neuronowe. Jeżeli chodzi o funkcję celu to będzie ona łączyć zarówno dokładność jak i precyzję w taki sposób, aby uniknąć potencjalnych nieoptymalnych rozwiązań.

- Klasyfikacja na danych testowych

Modele po wstępnym przetrenowaniu zostaną przyłożone do zbioru danych testowych w taki sposób, aby można było ocenić ich skuteczność w kontekście oceny czy dana osoba jest pełnoletnia czy też nie.

- Ocena wyniku

Ocena wyniku to krytyczny etap, w którym analizowane są rezultaty uzyskane na danych testowych. Na tym etapie porównuje się rzeczywisty wiek z przewidywanym przez model. Kluczowe miary, takie jak dokładność czy precyzja, pozwalają ocenić, jak bliskie rzeczywistym wartościom są przewidywania modelu. Wyniki te mogą także służyć do identyfikacji potencjalnych obszarów do dalszej optymalizacji modelu.

- Zapisanie modelu

Po zakończeniu treningu oraz ewaluacji, model jest zapisywany w formie umożliwiającej jego późniejsze użycie. Najczęściej model jest eksportowany do formatu umożliwiającego

jego łatwe wdrożenie, na przykład w formie pliku Pickle (Python) lub HDF5. Zapisanie modelu obejmuje również zapisanie jego architektury oraz wag, co umożliwia jego szybkie wczytanie i użycie w środowisku produkcyjnym. Rozbudowanie i zastosowanie opisanej metodologii pozwala na stworzenie efektywnego systemu do automatycznej estymacji wieku ze zdjęć, który może znaleźć zastosowanie w różnych dziedzinach.

Rozdział 3 Budowa modelu i ocena skuteczności

3.1 Przygotowanie danych oraz wstępna analiza

Budowanie każdego modelu uczenia maszynowego wymaga dokładnego przeanalizowania danych wejściowych. Takie podejście gwarantuje dokładne zrozumienie próbki treningowej oraz potencjalnie uchroni przed błędnym wnioskowaniem. Na podstawie zbioru danych etykiet zawartych w plikach zbioru UTK faces, została przeprowadzona ogólna analiza wieku, płci oraz etniczności osób znajdujących się na zdjęciach. Te dwie ostatnie kategorie, pozornie nieznaczące mogą dostarczyć ważnych informacji w kontekście późniejszej augmentacji danych, tak, aby uniknąć problemu związanego z brakiem reprezentatywności w poszczególnych grupach.

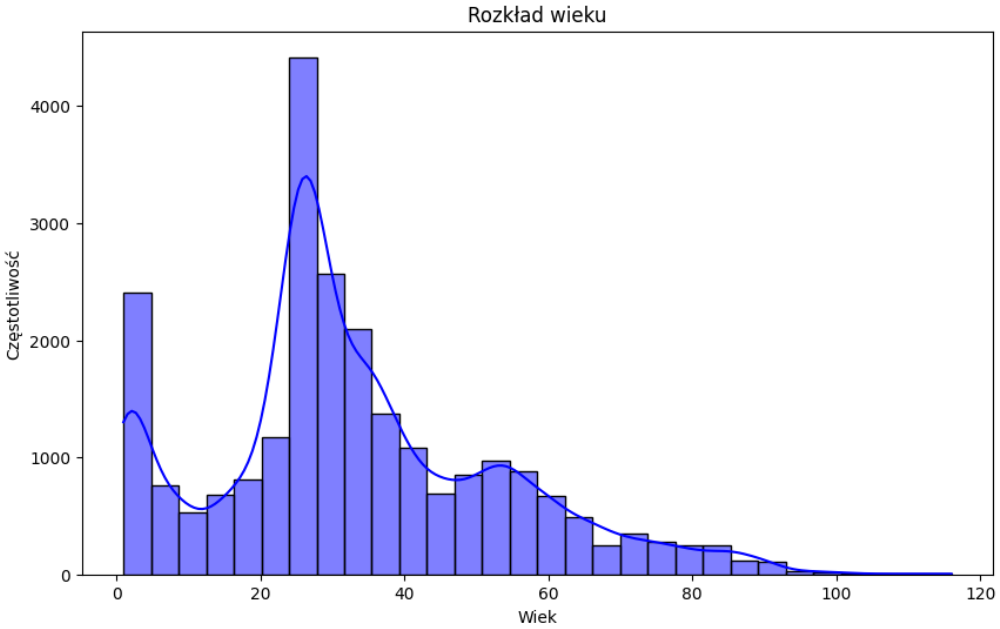
Tabela statystyczna podsumowuje rozkład wieku w zbiorze danych, przedstawiając takie wartości jak średnia (33 lata), mediana (29 lat), oraz zakres (od 1 do 116 lat). Takie zestawienie pomaga zrozumieć, że dane są dość zróżnicowane pod względem wieku, co jest korzystne dla treningu modelu predykcji wieku. Jednakże, ze względu na koncentrację próbek w przedziale wiekowym od 20 do 30 lat, model może być mniej dokładny w przewidywaniu wieku osób starszych lub bardzo młodych.

Tabela 2 Rozkład wieku ze zbioru danych UTK Face

Statystyka	Wiek
Ilość	24 102
Średnia	33,04
STD	20,13
Minimum	1
25 percentyl	23
Mediana	29
75 percentyl	45
Maksimum	116

Źródło: Opracowanie własne na podstawie danych ze zbioru UTK Face

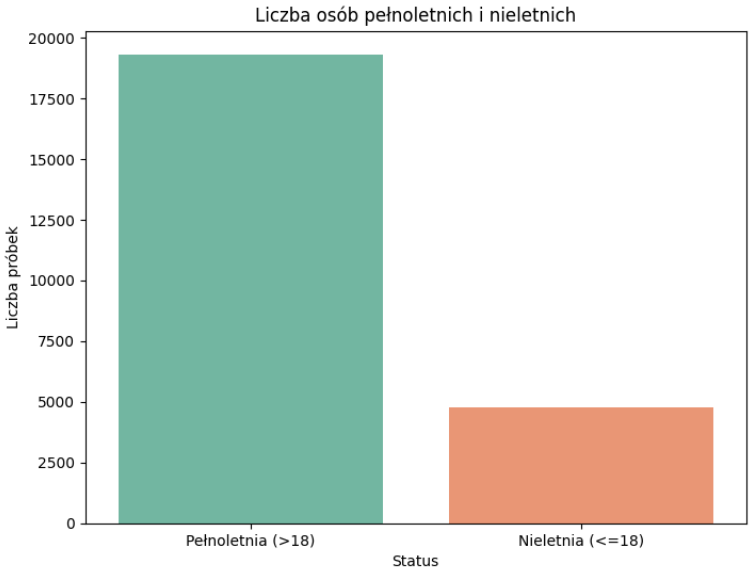
Rysunek 5 Rozkład wieku graficznie na podstawie zbioru UTK Face



Źródło: Opracowanie własne na podstawie danych ze zbioru UTK Face

Wykres histogramu dokładniej przedstawia rozkład wieku w całym zbiorze danych. Widoczny jest wyraźny szczyt w wieku około 20 lat, co sugeruje, że większość osób na zdjęciach jest w młodym wieku dorosłym. Jest to istotne dla modeli predykcyjnych, ponieważ może to wpłynąć na zdolność modelu do dokładnej klasyfikacji wieku osób starszych, które są mniej licznie reprezentowane. Obserwuje się również mniejsze szczyty w okolicach wieku dziecięcego (około 5 lat) oraz średniego wieku (około 40 lat).

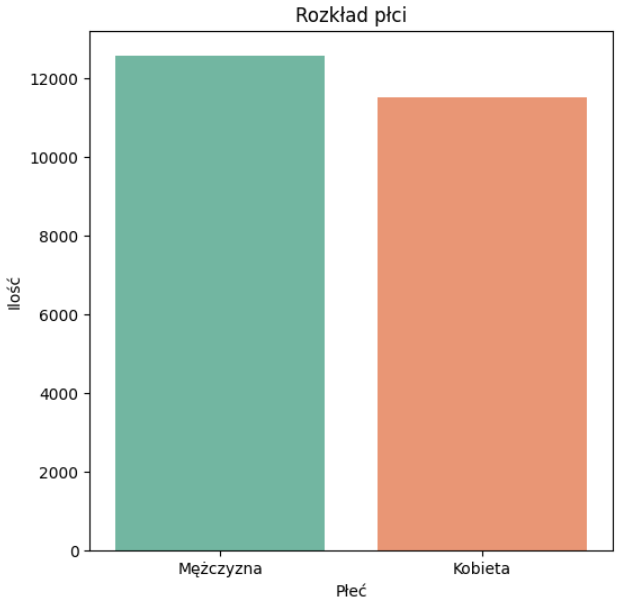
Rysunek 6 Liczna osób pełnoletnich oraz niepełnoletnich



Źródło: Opracowanie własne na podstawie danych ze zbioru UTK Face

Wykres słupkowy pokazuje liczbę osób pełnoletnich (powyżej 18 lat) oraz nieletnich (18 lat i młodsze). Zdecydowana większość próbek to osoby pełnoletnie, co może sugerować, że model będzie lepiej przystosowany do klasyfikacji wieku dorosłych niż dzieci. Przy projektowaniu modeli do predykcji wieku, ta dysproporcja powinna być brana pod uwagę, aby uniknąć stronniczości.

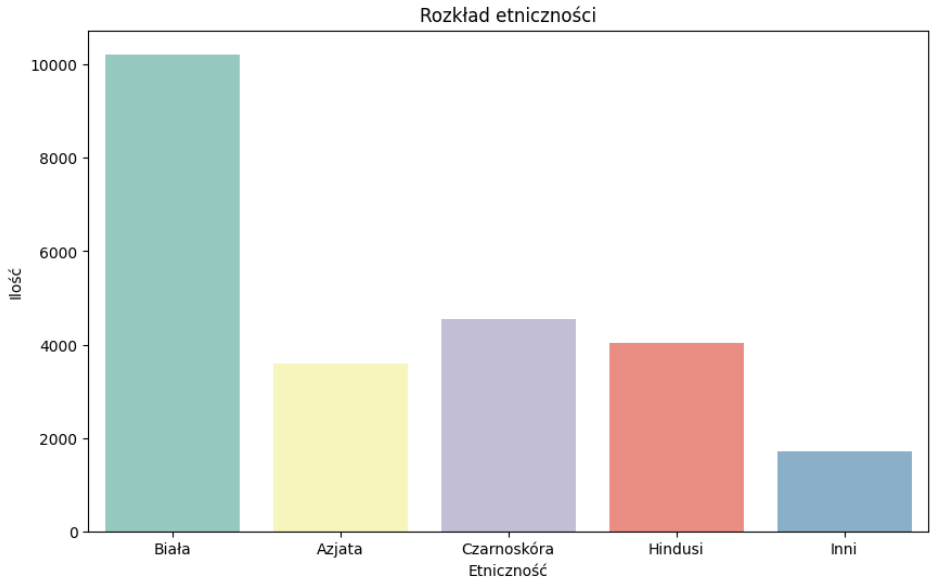
Rysunek 7 Podział zbioru danych w kategoriach płci



Źródło: Opracowanie własne na podstawie danych ze zbioru UTK Face

Wykres słupkowy przedstawia rozkład płci w zbiorze danych. Widać, że liczba mężczyzn i kobiet jest zbliżona, co jest korzystne dla modelu predykcyjnego, gdyż zminimalizuje ryzyko stronniczości względem jednej z płci. Równowaga w danych wejściowych jest kluczowa, zwłaszcza przy ocenie cech takich jak wiek, które mogą być różnie rozkładane w zależności od płci.

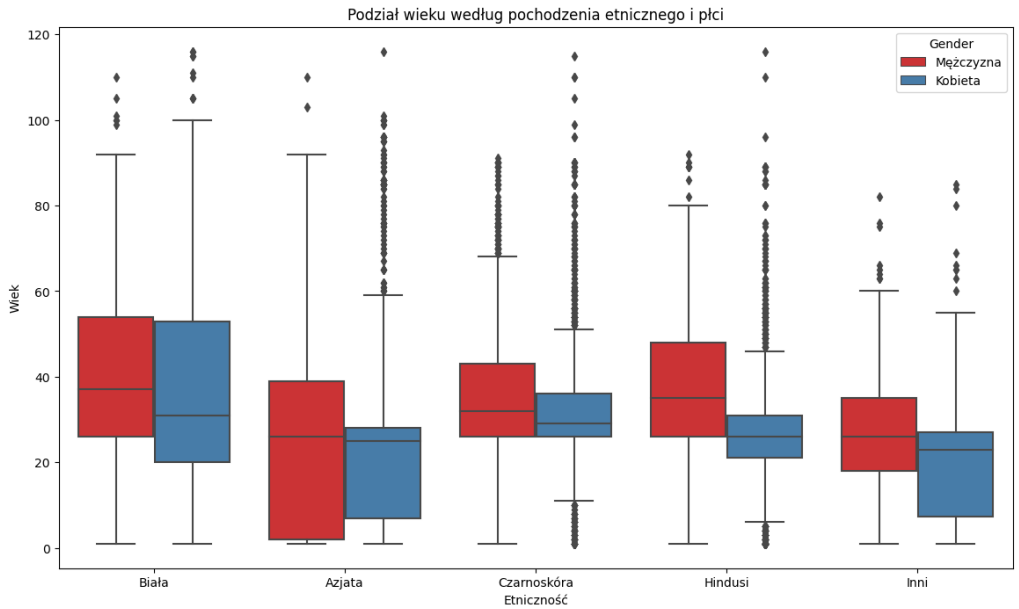
Rysunek 8 Rozkład etniczności



Źródło: Opracowanie własne na podstawie danych ze zbioru UTK Face

Ten wykres słupkowy przedstawia rozkład osób różnych grup etnicznych w zbiorze danych. Największą grupą są osoby rasy białej, co może mieć wpływ na zdolność modelu do generalizacji na inne grupy etniczne. Aby model był sprawiedliwy i dokładny, powinien dobrze działać niezależnie od etniczności osoby na zdjęciu, dlatego istotne jest, aby uwzględnić tę różnorodność podczas treningu.

Rysunek 9 Rozkład wieku według pochodzenia etnicznego i płci



Źródło: Opracowanie własne na podstawie danych ze zbioru UTK Face

Boxplot przedstawia rozkład wieku w różnych grupach etnicznych, z podziałem na płeć. Można zauważyć, że rozkłady wieku różnią się między grupami etnicznymi, a także pomiędzy płciami w obrębie tych grup. W przypadku oceny wieku ze zdjęć, taki podział pozwala na zrozumienie, czy model może potrzebować uwzględnić te różnice, aby dokładnie przewidywać wiek w różnych podgrupach demograficznych.

Podsumowując, rozkłady wieku, płci i etniczności w zbiorze danych UTK Faces wskazują na pewne wyzwania, takie jak dominacja próbek dorosłych i osób rasy białej, które mogą wpłynąć na wyniki modeli oceniających wiek na podstawie zdjęć. Ważne jest, aby uwzględnić te czynniki przy trenowaniu modelu, aby osiągnąć jak najbardziej dokładne i sprawiedliwe rezultaty.

Kolejnym etapem będzie przygotowanie danych do finalnego modelowania z użyciem głębokich sieci neuronowych w tym celu zostały przeprowadzone następujące kroki:

- Inicjalizacja detektora twarzy: Inicjalizowany jest detektor twarzy oparty na kaskadach Haar, który będzie wykorzystywany do wykrywania twarzy na każdym zdjęciu.
- Augmentacja: Augmentacja obejmuje taki możliwości jak Odbicie w poziomie z prawdopodobieństwem 50%, Losowy obrót w zakresie -20 do 20 stopni oraz skalowanie w zakresie 80% do 120%, Zmianę jasności obrazu oraz dodanie szumu Gaussa. Augmentacja została zastosowana dla zdjęć osób, które są nieletnie ze względu na zwiększenie próbki.
- Przygotowanie obrazu: Wycięta twarz jest przeskalowywana do rozmiaru 128x128 pikseli.
- Przygotowanie etykiet: Z każdego zdjęcia brana jest informacja na temat wieku danej osoby, która później jest kategoryzowana na osoby niepełnoletnie oraz pełnoletnie.

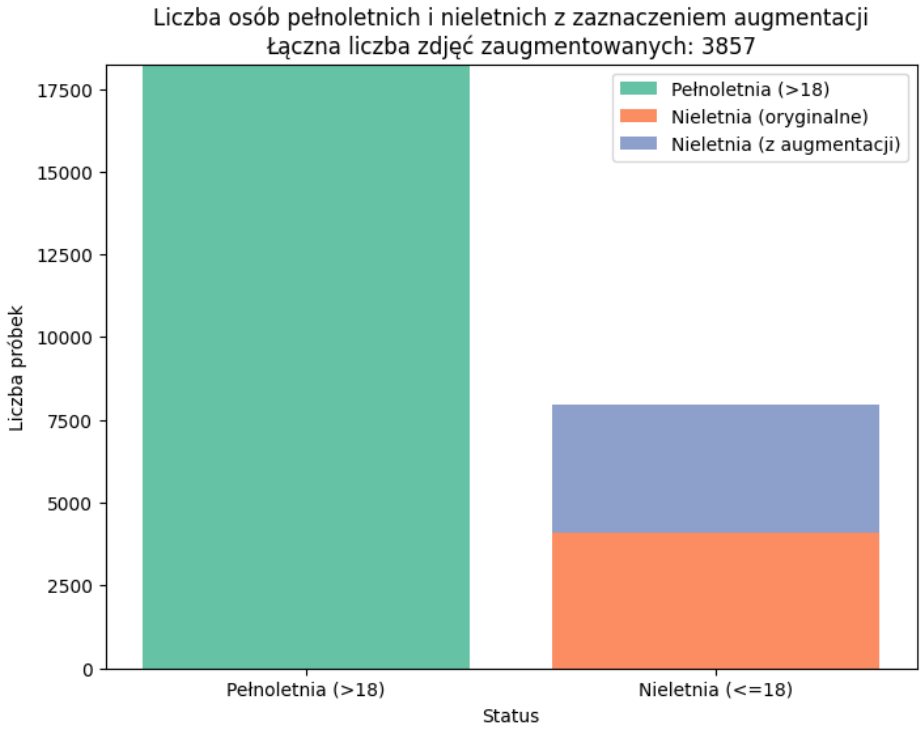
Finalny wynik procesu przygotowywania danych został przedstawiony na rysunku poniżej.

Rysunek 10 Przykład finalnie przygotowanych danych

Źródło: Opracowanie własne na podstawie danych ze zbioru UTK Face

Na rysunku numer 10 można zauważyć 6 przykładowych zdjęć z wcześniej przygotowanego zbioru danych. Zdjęcie niepełnoletniej osoby w środkowym rzędzie zostało wygenerowane na podstawie zdjęcie w lewym górnym rzędzie. Finalnie do zbioru danych dodatkowo zostało dodanych 3875 zdjęć, które zostały augmentowane na podstawie zdjęć osób nieletnich. Finalny rozkład osób pełnoletnich oraz osób nieletnich został zaprezentowany na rysunku numer 11.

Rysunek 11 Liczba osób pełnoletnich oraz nieletnich z zaznaczeniem augmentacji



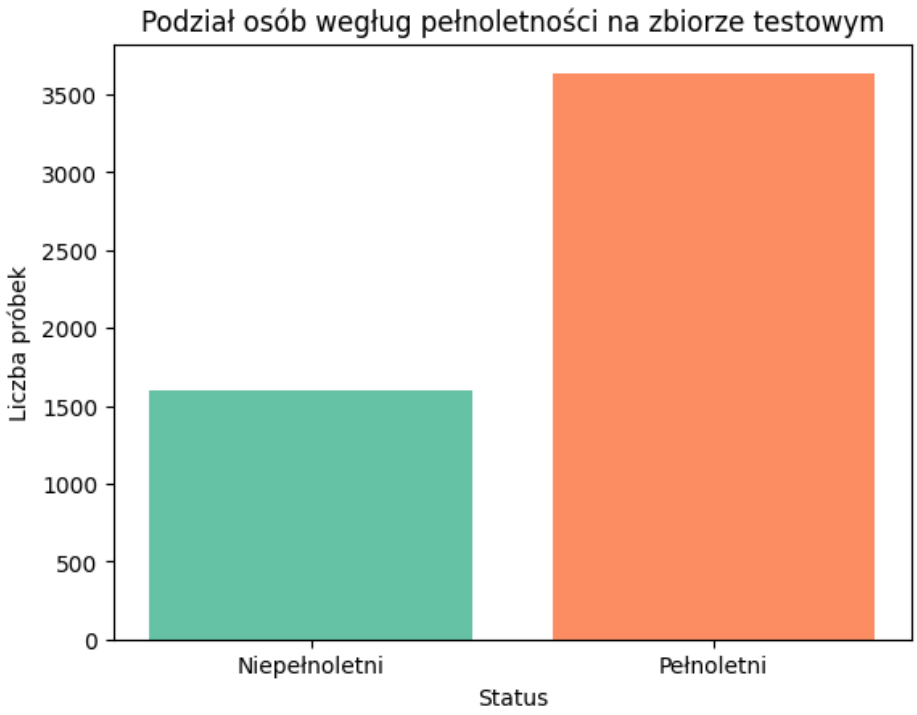
Źródło: Opracowanie własne na podstawie danych ze zbioru UTK Face

3.2 Zastosowanie modelu

Po wcześniejszym przygotowaniu danych kolejną częścią będzie wybranie odpowiedniej próbki testowej, treningowej oraz walidacyjnej a także wybranie oraz wytrenowanie modelu. Projektowanie własnej architektury konwolucyjnej sieci neuronowej (CNN) do rozpoznawania pełnoletności na podstawie zdjęcia to zadanie, które wymaga starannego dobierania warstw i ich parametrów, aby osiągnąć optymalną wydajność. Rysunek numer 12 przedstawia strukturę finalnego modelu opartego na konwolucyjnych sieciach neuronowych, który został wstępnie wytrenowany na zbiorze danych treningowych. Dane treningowe zostały ustalone na 80 % ogólnego zbioru danych natomiast dane testowe stanowią 20% losowo wybranych zdjęć. Zbiór walidacyjny, jaki został zastosowany w tym modelu jest taki sam jak zbiór danych testowych, czyli stanowi losowo wybrane 20 % ogólnego zbioru danych nieuczestniczących w procesie uczenia. Metoda losowania zastosowana w tym przypadku to losowe próbkowanie bez zwracania. Oznacza to, że każdy przykład z oryginalnego zbioru danych ma jednakową szansę znalezienia się w zbiorze treningowym lub testowym, a raz wybrany przykład nie może zostać wybrany ponownie do

innego zbioru. Takie losowanie zapewnia, że próbki są unikalne w każdym podzbiorze i reprezentatywne dla całego zbioru danych. Na wykresie poniżej został zaprezentowany rozkład finalnej etykiety użytej do modelowania.

Rysunek 11 Liczba osób pełnoletnich oraz nieletnich na zbiorze testowym oraz walidacyjnym



Źródło: Opracowanie własne na podstawie danych ze zbioru UTK Face

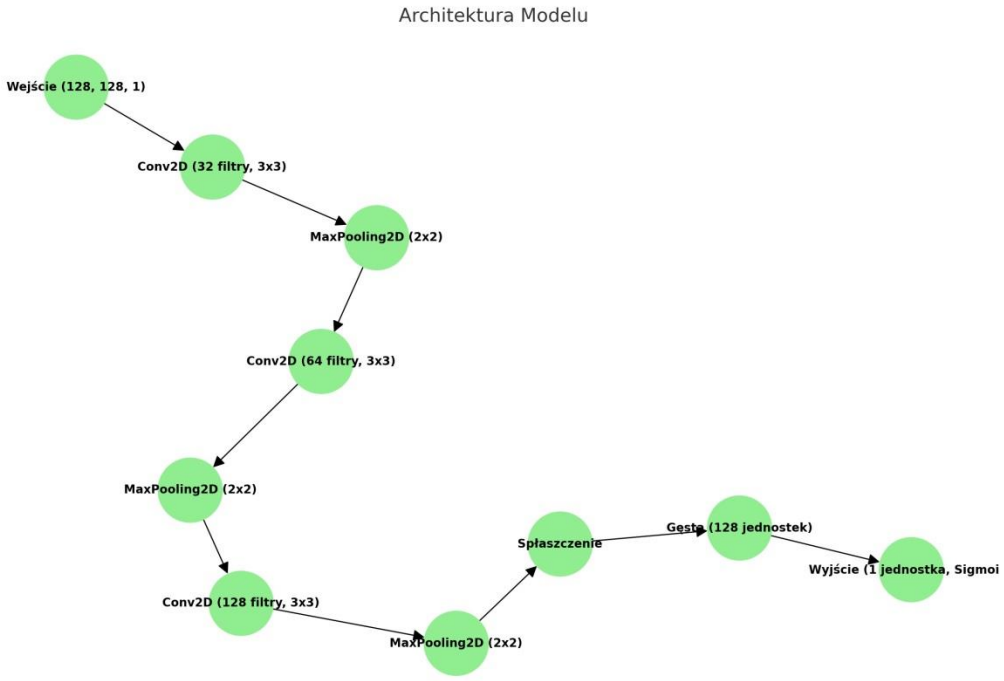
Analiza tego wykresu pozwala stwierdzić, że proporcja osób pełnoletnich a niepełnoletnich w zbiorze testowym jest zachowana, co sprawia, że próba testowa jest reprezentatywna w stosunku do zbioru treningowego.

Zbiór walidacyjny odgrywa kluczową rolę w procesie trenowania modelu. Jest to podzbiór danych, który nie jest używany bezpośrednio do uczenia modelu, ale służy do monitorowania jego wydajności podczas treningu. Dzięki zbiorowi walidacyjnemu można na bieżąco oceniać, jak dobrze model generalizuje na danych niewidzianych podczas treningu, co pozwala na wykrywanie nadmiernego dopasowania (overfittingu). Dodatkowo Zbiór walidacyjny umożliwia eksperymentowanie z różnymi ustawieniami hiperparametrów, takich jak współczynnik uczenia się czy liczba warstw lub filtrów, bez ryzyka wpływu na ostateczną ocenę modelu.

Proces walidacji został zrealizowany poprzez zastosowanie metody walidacji z wydzielonym zbiorem. Polega ona na jednokrotnym podziale danych na dwa zbiory: treningowy i walidacyjny. W przeciwieństwie do metod takich jak walidacja krzyżowa, gdzie

dane są wielokrotnie dzielone na różne podzbiory, walidacja z wydzielonym zbiorem opiera się na jednorazowym losowym podziale. Model jest trenowany na zbiorze treningowym, a po każdej epoce jego wydajność jest oceniana na zbiorze walidacyjnym. Taki sposób walidacji pozwala na monitorowanie zdolności modelu do generalizacji, wykrywanie problemów z przeuczeniem oraz optymalizację procesu uczenia poprzez dostosowanie hiperparametrów.

Rysunek 12 Architektura konwolucyjnych sieci neuronowych



Źródło: Opracowanie własne

Tak przedstawiony model możemy rozłożyć na poszczególne warstwy:

- Warstwa konwolucyjna Conv2D – Warstwy konwolucyjne są fundamentem modeli CNN i służą do wyodrębniania cech z obrazów. W zadaniu estymacji wieku, te cechy mogą obejmować kontury twarzy, zmarszczki, kształt oczu oraz ust. Pierwsza warstwa konwolucyjna używa 32 filtrów o rozmiarze 3x3, aby wyodrębnić podstawowe cechy z obrazu wejściowego, takie jak krawędzie, kąty itp. Funkcja aktywacji ReLU (Rectified Linear Unit) jest używana, aby wprowadzić nieliniowość do procesu uczenia, pomagając modelowi lepiej radzić sobie z złożonymi wzorcami w danych.
- Warstwa MaxPooling2D – Pooling redukuje wymiarowość każdej cechy przy jednoczesnym zachowaniu najważniejszych informacji. Pomaga to w zmniejszeniu liczby parametrów i obliczeń, co zwiększa efektywność treningu i zapobiega przeuczeniu. Ta warstwa służy do zmniejszenia wymiarów przestrzennych obrazu.

Wybiera maksymalną wartość z każdego kwadratu 2x2 pikseli, co pomaga w redukcji liczby parametrów i obliczeń w sieci oraz zapobiega przeuczeniu (overfitting).

- Drugie warstwy konwolucyjne i poolingowe – Druga warstwa konwolucyjna zwiększa liczbę filtrów, do 64, co pozwala na wykrywanie bardziej złożonych cech, jak tekstury i większe kształty. Kolejna warstwa poolingowa dalej redukuje wymiary obrazu, zachowując przy tym najważniejsze cechy.
- Trzecia warstwa konwolucyjna i poolingowa – Trzecia warstwa konwolucyjna jeszcze bardziej zwiększa liczbę filtrów, umożliwiając modelowi wykrywanie jeszcze bardziej złożonych i abstrakcyjnych cech. Ostatnia warstwa poolingowa redukuje przestrzenne wymiary danych wyjściowych, minimalizując ryzyko przeuczenia.
- Warstwa Flatten – Spłaszcza wielowymiarowe dane wyjściowe z poprzedniej warstwy do jednowymiarowego wektora. Jest to niezbędne, ponieważ kolejne w pełni połączone warstwy (Dense) oczekują jednowymiarowego wektora, jako wejścia.
- Warstwa Dense – Warstwy Dense wykorzystują wyodrębnione cechy do przeprowadzenia klasyfikacji. Łączona warstwa sieci z 128 jednostkami, które uczą się kombinacji cech z poprzednich warstw, które są najbardziej informatywne dla przewidywania wieku.
- Wyjściowa warstwa Dense – Ostatnia warstwa modelu używa funkcji aktywacji sigmoid, aby wygenerować pojedynczą wartość między 0 a 1, co jest interpretowane, jako prawdopodobieństwo przynależności do jednej z klas (np. osoby powyżej 18 lat).

Funkcja celu, jaka została zastosowana w tym modelu to połączenie 3 funkcji bazujących na dokładności klasyfikacji: dokładność, precyzja oraz wrażliwość. Takie połączenie funkcji celu będzie gwarantować dokładne oszacowania przy zachowaniu odpowiednich proporcji predykcji 0 oraz 1. Tabela numer 3 przedstawia wyniki trenowania modelu w podziale na Epoki, czyli arbitralne granice ogólnie definiowana, jako „jedno przejście przez cały zbiór danych”, używana do rozdzielania szkolenia na odrębne fazy.

Tabela 3 Przeprowadzenie treningu konwolucyjnych sieci neuronowych na danych treningowych

Epoch	Loss	Accuracy	Precision	Recall	Val_Loss	Val_Accuracy	Val_Precision	Val_Recall
1	1.0649	0.8644	0.8740	0.9410	0.2932	0.8807	0.9323	0.8931
2	0.2849	0.8956	0.8995	0.9571	0.2679	0.9033	0.8997	0.9687
3	0.2579	0.9062	0.9063	0.9651	0.2609	0.9103	0.9114	0.9645
4	0.2250	0.9207	0.9168	0.9747	0.7892	0.8329	0.8068	0.9984
5	0.2052	0.9253	0.9226	0.9747	0.2669	0.9052	0.9246	0.9401
6	0.2076	0.9257	0.9232	0.9745	0.2757	0.9109	0.9096	0.9678
7	0.1650	0.9403	0.9364	0.9809	0.2434	0.9168	0.9116	0.9747
8	0.1483	0.9486	0.9451	0.9834	0.2639	0.9185	0.9204	0.9662
9	0.1316	0.9527	0.9493	0.9848	0.2922	0.9212	0.9081	0.9863
10	0.1174	0.9563	0.9542	0.9845	0.4108	0.9145	0.9098	0.9733

- Trendy treningowe: Ogólnie, model wykazuje stabilny wzrost dokładności i precyzji z kolejnymi epokami, jednocześnie obniżając stratę, co świadczy o efektywności procesu nauki.
- Wyniki walidacji - Pomimo ogólnie dobrych wyników, model wykazuje niektóre oznaki możliwego przeuczenia, szczególnie widoczne w 4 epoce, gdzie strata walidacyjna gwałtownie wzrasta. Może to sugerować, że model zbytnio dopasowuje się do danych treningowych kosztem zdolności do generalizacji na nowych danych.
- Potrzeba dostosowania - Aby poprawić generalizację, można by rozważyć wprowadzenie technik regularyzacji, takich jak dropout lub regularyzacja L1/L2, a także dalsze dostosowanie parametrów modelu lub procedur wczesnego zatrzymywania.

Na podstawie tych wyników oraz wniosków do modelu zostały zastosowane poprawki w taki sposób, aby uniknąć problemu przeuczenia. Po pierwsze dodanie regularyzacji L2 do warstw konwolucyjnych i gęstych warstw w celu przeciwdziałania przeuczeniu. Pomaga to w redukcji nadmiernej złożoności modelu, co zmniejsza różnicę między stratą treningową a walidacyjną. W kolejnym kroku została dodana warstwa Dropout po ostatniej warstwie gęstej przed warstwą wyjściową. Dropout losowo wyłącza 50% neuronów podczas treningu, co zmniejsza ryzyko przeuczenia. Szybkość uczenia została zmniejszona do 0.0001 w optimizerze Adam. Powolniejsze tempo uczenia pozwala modelowi lepiej dostosować się do danych bez zbyt gwałtownych zmian w wagach, co może również poprawić stabilność treningu. Wyniki treningu poprawionego modelu zostały zaprezentowane na tabeli poniżej.

Tabela 4 Przeprowadzenie treningu konwolucyjnych sieci neuronowych na danych treningowych po poprawkach

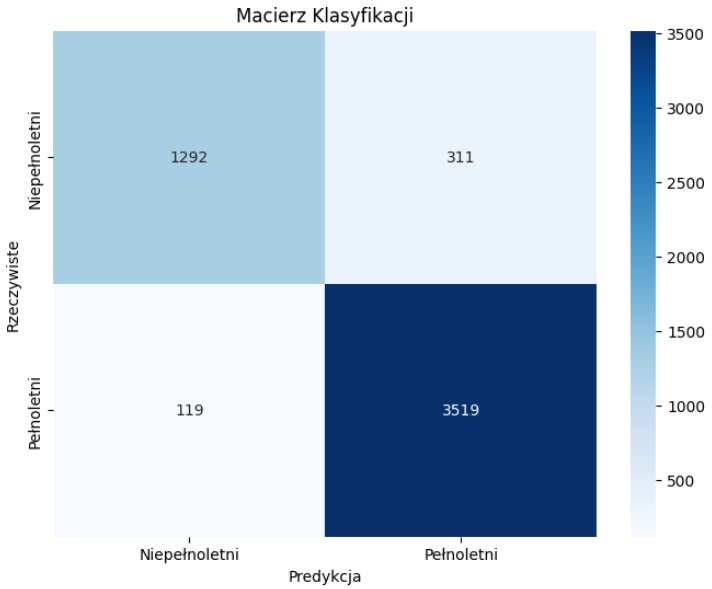
Epoch	Loss	Accuracy	Precision	Recall	Val_Loss	Val_Accuracy	Val_Precision	Val_Recall
1	0.8990	0.7836	0.7934	0.9324	0.5803	0.8601	0.8382	0.9896
2	0.5702	0.8613	0.8630	0.9520	0.4894	0.9000	0.8986	0.9648
3	0.5121	0.8842	0.8863	0.9566	0.4865	0.8926	0.9372	0.9060
4	0.4643	0.9001	0.8986	0.9656	0.4521	0.9021	0.9371	0.9208
5	0.4320	0.9074	0.9054	0.9683	0.3947	0.9218	0.9243	0.9665
6	0.4036	0.9171	0.9132	0.9736	0.4166	0.9073	0.8868	0.9931
7	0.3833	0.9206	0.9171	0.9741	0.3824	0.9199	0.9049	0.9885
8	0.3572	0.9252	0.9210	0.9765	0.3459	0.9273	0.9381	0.9585
9	0.3298	0.9345	0.9312	0.9784	0.3299	0.9279	0.9267	0.9731
10	0.3095	0.9380	0.9333	0.9811	0.3312	0.9262	0.9091	0.9929

Jak można zauważyć z każdą kolejną epoką widzimy spadek wartości „Loss” i „Val_Loss”, co wskazuje, że model poprawia swoje dopasowanie do danych treningowych i walidacyjnych. Dokładność, precyzja, i wrażliwość rosną, co oznacza, że model jest coraz bardziej dokładny, precyzyjny, i czuły. Strata walidacyjna zmniejsza się stopniowo, co sugeruje, że model dobrze generalizuje na zbiorze walidacyjnym. W ostatniej epoce wartość „Val_Loss” wynosi 0.3312, co jest bliskie wartości „Loss” (0.3095), co oznacza stabilność modelu. Model nie wykazuje znaczących oznak przeuczenia. Wartości „Val_Loss” pozostają stosunkowo niskie, a różnice między „Loss” a „Val_Loss” nie są duże. Wyniki na zbiorze walidacyjnym (val_accuracy, val_precision, val_recall) są zbliżone do wyników na zbiorze treningowym, co świadczy o stabilności modelu. Z tej analizy można wywnioskować, że model działa poprawnie i generalizuje dobrze, co sugeruje, że wprowadzone zmiany w architekturze poprawiły jego wydajność.

3.3 Ocena jakości

Po wytrenowaniu modelu na danych treningowych, kolejnym krokiem będzie oszacowanie, jakości modelu. W tym celu na zbiór danych testowych został oszacowany za pomocą modelu pierwszego przed poprawkami oraz modelu drugiego po poprawkach. Do oceny, jakości modelu została zastosowana macierz klasyfikacji oraz krzywa ROC. Wyniki zostały zaprezentowane na rysunkach poniżej.

Rysunek 13 Macierz klasyfikacji modelu pierwszego

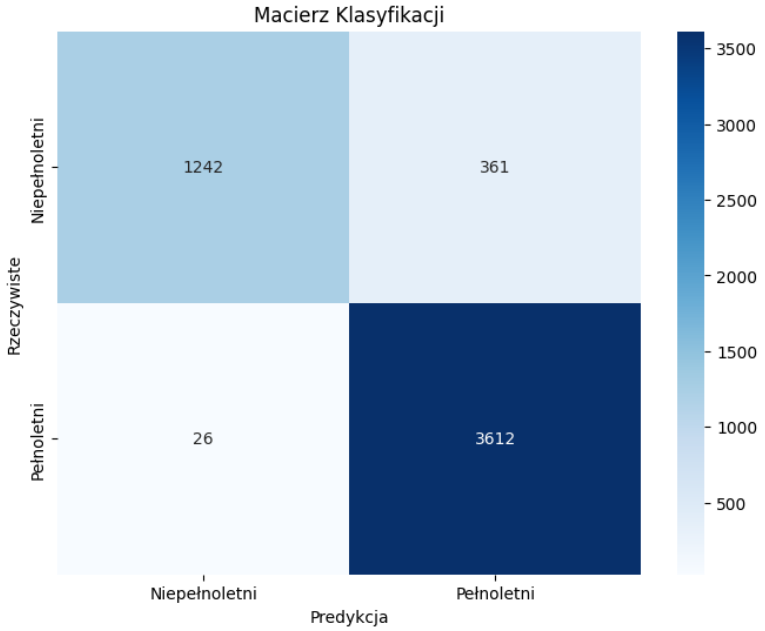


Macierz klasyfikacji przedstawia liczbę poprawnych i błędnych predykcji dokonanych przez model. W tym przypadku macierz została wykorzystana do oceny modelu na podstawie danych testowych.

- Niepełnoletni - Rzeczywiste vs Predykcja - 1292 to liczba osób, które są rzeczywiście niepełnoletnie i zostały poprawnie zaklasyfikowane, jako niepełnoletnie przez model. 311 to liczba osób, które są rzeczywiście niepełnoletnie, ale zostały błędnie zaklasyfikowane, jako pełnoletnie.
- Pełnoletni - Rzeczywiste vs Predykcja - 119 to liczba osób, które są rzeczywiście pełnoletnie, ale zostały błędnie zaklasyfikowane, jako niepełnoletnie. 3519 to liczba osób, które są rzeczywiście pełnoletnie i zostały poprawnie zaklasyfikowane, jako pełnoletnie przez model.

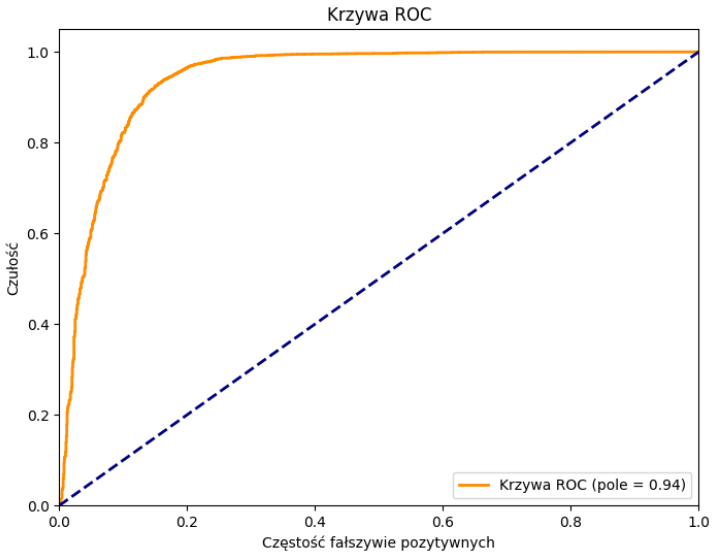
Model skutecznie identyfikuje osoby pełnoletnie, co jest widoczne w dużej liczbie prawidłowych klasyfikacji w dolnym prawym kwadracie. Błędy klasyfikacji są stosunkowo niskie, jednak model częściej popełnia błędy w przewidywaniu, czy osoby niepełnoletnie są pełnoletnie, niż odwrotnie.

Rysunek 14 Macierz klasyfikacji modelu drugiego



Model po dokonaniu modyfikacji osiągnął delikatnie większą dokładność na zbiorze testowym, widać tutaj zdecydowanie mniej przypadków fałszywie pozytywnych, natomiast delikatnie zwiększyła się ilość przypadków fałszywie negatywnych, czyli sytuacji, w których osoba niepełnoletnia jest klasyfikowana, jako osoba pełnoletnia.

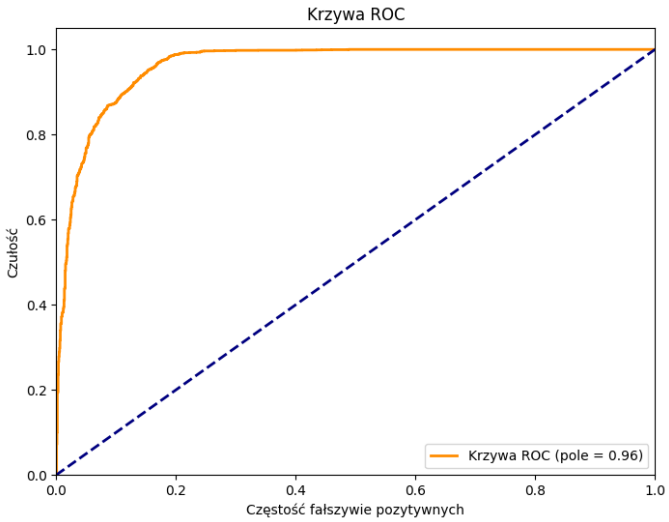
Rysunek 15 Wykres krzywej ROC modelu pierwszego



Krzywa ROC (ang. Receiver Operating Characteristic) służy do oceny zdolności klasyfikacyjnej modelu. Przedstawia zależność między czułością (True Positive Rate) a częstością fałszywych pozytywnych (False Positive Rate) dla różnych progów decyzyjnych. Wysokie AUC (0.94) wskazuje, że model jest bardzo skuteczny w rozpoznawaniu

pełnoletności na podstawie zdjęć. Model radzi sobie dobrze z minimalizacją zarówno fałszywych alarmów (błędnie klasyfikowane niepełnoletnie osoby) jak i przegapianiem rzeczywiście pełnoletnich.

Rysunek 16 Wykres krzywej ROC modelu drugiego



Wartości na drugim wykresie krzywej ROC są niemal identyczne jak w przypadku pierwszego modelu. Widać tutaj jedynie większą dokładność (0,96).

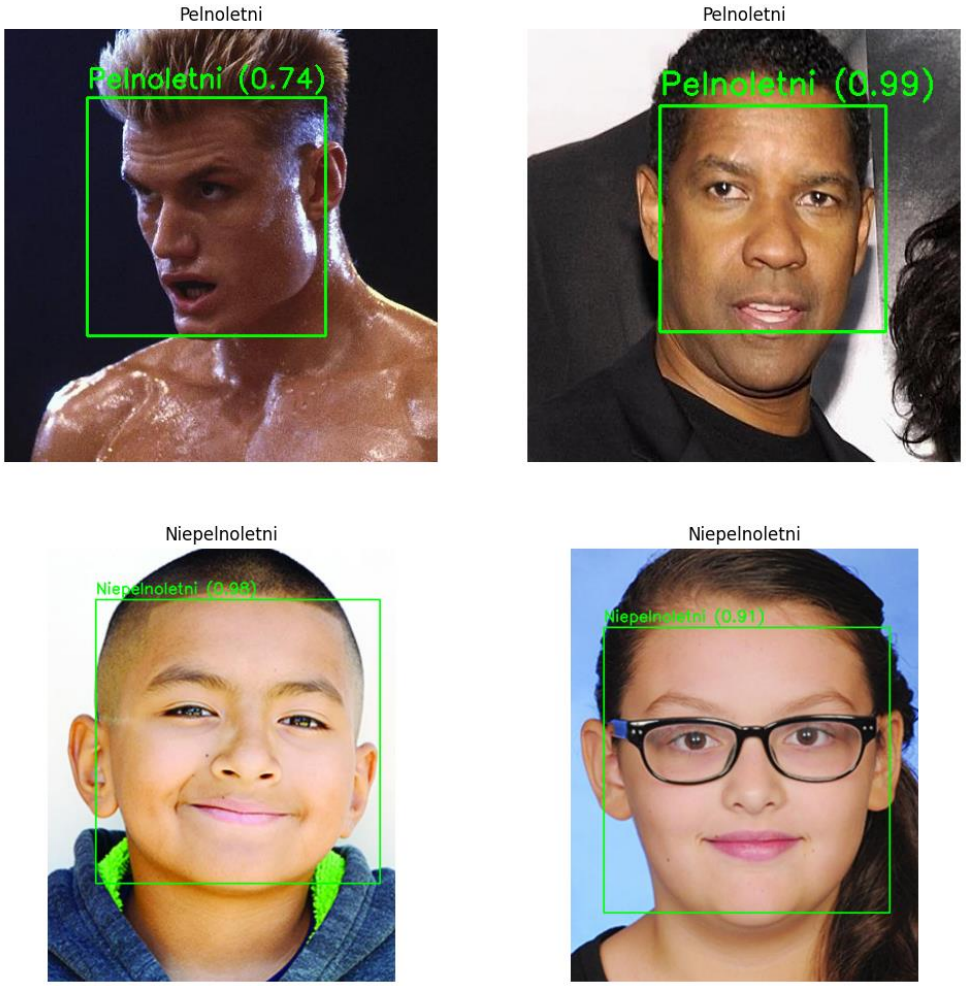
3.4 Implementacja oraz przykłady użycia

Finalnie wytrenowany model możemy wykorzystać na wiele sposobów. Po zakończeniu procesu trenowania modelu konwolucyjnej sieci neuronowej (CNN) w celu przewidywania pełnoletności na podstawie zdjęcia twarzy, kluczowe jest zapisanie tego modelu w sposób umożliwiający jego późniejsze wykorzystanie. W tym celu można skorzystać z funkcji biblioteki TensorFlow, która zapisuje całą strukturę modelu, jego wagi oraz konfigurację treningową w formacie HDF5. Dzięki temu model może być łatwo załadowany i użyty w różnych środowiskach, w tym na stronach internetowych czy też lokalnym innym środowisku.

Aby wykorzystać zapisany model na stronie internetowej, należy go najpierw załadować przy użyciu funkcji „load_model”. Gdy użytkownik przesyła swoje zdjęcie na stronie, obraz jest przetwarzany w taki sposób, aby był gotowy do wprowadzenia do modelu. Co ważne zdjęcie musi zostać przygotowane w taki sam sposób, co dane treningowe wykorzystane w modelu. Tak przygotowany obraz jest następnie przesyłany do modelu za pomocą funkcji „prediction”, co zwraca prawdopodobieństwo, że osoba na zdjęciu jest pełnoletnia. Wynik tej predykcji jest wyświetlany na stronie w formie etykiety, która informuje, czy dana osoba jest

pełnoletnia, z przypisanym poziomem pewności. Na przykład, jak przedstawiono na rysunku 17, powyżej twarzy użytkownika pojawia się informacja "Pełnoletni" lub "Niepełnoletni", w zależności od wyniku predykcji. Taki system może być używany w aplikacjach wymagających automatycznej weryfikacji wieku, jak np. zakupy online, gdzie konieczne jest potwierdzenie pełnoletności przed dokonaniem transakcji.

Rysunek 17 Przykłady użycia modelu na innych zdjęciach



Dzięki zastosowaniu wytrenowanego modelu CNN, proces weryfikacji wieku staje się szybki, automatyczny, i bardziej niezawodny, eliminując konieczność ręcznej kontroli dokumentów tożsamości przez personel. To rozwiązanie jest szczególnie korzystne w kontekście handlu detalicznego i systemów bezpieczeństwa, gdzie czas i dokładność mają kluczowe znaczenie.

Podsumowanie

Przeprowadzone badania oraz uzyskane wyniki potwierdziły potencjał konwolucyjnych sieci neuronowych (CNN) w dziedzinie automatycznej oceny wieku na podstawie obrazów twarzy. Rozwój technologii głębokiego uczenia umożliwia coraz bardziej precyzyjną analizę obrazów, co ma szerokie zastosowanie w różnych dziedzinach, takich jak medycyna, handel detaliczny czy też bezpieczeństwo na stronach internetowych. W toku pracy udało się zrealizować założone cele, począwszy od przeglądu literatury oraz dostępnych zbiorów danych, poprzez implementację i optymalizację modelu CNN, aż po szczegółową analizę uzyskanych wyników. Wybór odpowiedniej architektury modelu oraz technik przetwarzania danych okazał się kluczowy dla uzyskania wysokiej skuteczności modelu. Należy jednak podkreślić, że pomimo uzyskanych sukcesów, istnieje szereg wyzwań, które powinny być adresowane w przyszłych badaniach. Jednym z kluczowych problemów jest zapewnienie odpowiedniej, jakości i reprezentatywności danych treningowych, co może mieć istotny wpływ na ogólną skuteczność systemu. Ponadto, istotne jest rozwijanie algorytmów uwzględniających zmienność warunków oświetleniowych czy też kąta ustawienia twarzy. Warto podkreślić, że potencjalne kierunki przyszłych badań powinny koncentrować się na dalszym doskonaleniu algorytmów głębokiego uczenia, z uwzględnieniem szerszej gamy czynników wpływających na proces starzenia się twarzy. Otwiera to nowe możliwości dla rozwoju narzędzi do oceny wieku, które mogą przyczynić się do postępu w różnych dziedzinach życia społecznego i gospodarczego. Ostatecznie, niniejsza praca nie tylko przyczyniła się do stworzenia efektywnego narzędzia do oceny wieku na podstawie zdjęć, ale również zwróciła uwagę na potencjalne korzyści i wyzwania związane z wdrażaniem tej technologii w praktyce. Jej wyniki stanowią solidną podstawę do dalszych badań oraz wdrożeń praktycznych w przyszłości.

Bibliografia

Al-Shannaq, Arwa S., Elrefaei, L. *„Comprehensive Analysis of The Literature for Age Estimation from Facial Images”* IEEE Access (2019)

Angulu, R., Tapamo, J.R. & Adewumi, A.O. *„Age estimation via face images: a survey”. J Image Video Proc. 2018, 42* (2018)

Coleman, S. R., & Grover, R. *„The anatomy of the aging face: Volume loss and changes in 3-dimensional topography.”* Aesthetic Surgery Journal (2006)

Dokumentacja biblioteki Face Recognition dostępna: 20 Sierpnia 2024 < https://face-recognition.readthedocs.io/en/latest/readme.html#>

Dokumentacja biblioteki MTCNN dostępna na: < https://github.com/ipazc/mtcnn?tab=readme-ov-file#zhang2016>

Dokumentacja biblioteki OpenCV Dostęp: 20 Sierpnia 2024: <https://docs.opencv.org/4.x/index.html>

Dokumentacja biblioteki TensorFlow Dostęp: 20 Sierpnia 2024 < https://github.com/tensorflow/docs>

Elkarazle, K. *„Facial Age Classification Using Deep Learning and Generative Adversarial Networks”* (2022)

ELKarazle, K., Raman, V., Then, P. *„Facial Age Estimation Using Machine Learning Techniques: An Overview”* Big Data Cogn. Comput. (2022)

Geng, X., C., Yin, Z. -H. Zhou, *"Facial Age Estimation by Learning from Label Distributions"* IEEE Transactions on Pattern Analysis and Machine Intelligence (2013)

Ghrban, Z., El abbadi, N. *„Gender and Age Estimation from Human Faces Based on Deep Learning Techniques: A Review”* International Journal of Computing and Digital Systems (2023)

Girasa, R. *„Ethics and Privacy I: Facial Recognition and Robotics”* Artificial Intelligence as a Disruptive Technology (2020)

Greco, A. *„Guess the Age 2021: Age Estimation from Facial Images with Deep Convolutional Neural Networks”* Computer Analysis of Images and Patterns (2021)

Greco, A., Saggese, A., Vento, M., Vigilante, V., *„Effective training of convolutional neural networks for age estimation based on knowledge distillation”* Neural Computing and Applications (2022)

Han, H., Otto, C., & Jain, A. K. *„Age estimation from face images: Human vs. machine performance.”* 6th IAPR International Conference on Biometrics (ICB) (2013).

Kanan, C., Cotterell G., W. *„Color-to-Grayscale: Does the Method Matter in Image Recognition?”* Plos One (2012)

Kjaerran, A., Stray Bugge, E., Bakke Vennerod C., *„Facial Age Estimation Using Convolutional Neural Networks”* (2021)

Kuang C., Huang, Z,, Zhang, W. *„Deeply Learned Rich Coding for Cross-Dataset Facial Age Estimation”* IEEE International Conference on Computer Vision Workshop (2015)

Marques, I. *„Face Recognition Algorithms”* (2010)

Parisa Beham, M., Mohamed Mansoor Roomi, S. *„Face Recognition Using Appearance Based Approach: A Literature Survey”* International Journal of Computer Applications (IJCA) (2012)

Rahman, M.A., Aonty, S.S., Deb, K., Sarker, I.H. *„Attention-Based Human Age Estimation from Face Images to Enhance Public Security”* Data (2023)

Rosebrock, A. *„Face detection with dlib (HOG and CNN),,* Dostęp: 20 Sierpnia 2024 < *https*://pyimagesearch.com/2021/04/19/face-detection-with-dlib-hog-and-cnn/> (2021)

Saponara, S., Elhanashi, A. *„Impact of Image Resizing on Deep Learning Detectors for Training Time and Model Performance”* Applications in Electronics Pervading Industry (2022)

Shorten, C., Khoshgoftaar, T.M. *„A survey on Image Data Augmentation for Deep Learning”* J Big Data (2019)

Tan Yeh Ping, S., Hui Weng, C., Lau, B. *„Face detection through template matching and color segmentation”* EE 368 Final Project (2016)

Torrey, L., Shavlik, J., „Transfer *Learning”* (2019)

Turner Lee, N., Chin-Rothmann, C., *„Police surveillance and facial recognition: Why data privacy is imperative for communities of color”* Brookings (2022) Dostęp: 20 Sierpnia 2024 < https://www.brookings.edu/articles/police-surveillance-and-facial-recognition-why-data-privacy-is-an-imperative-for-communities-of-color/>

Wagner, P. *„Face Recognition with Python”* (2012) Dostęp: 20 Sierpnia 2024 < https://www.bytefish.de>

Yang, Ming-Hsuan & Kriegman, David & Ahuja, Narendra. *„Detecting Faces in Images: A Survey”* Pattern Analysis and Machine Intelligence IEEE Transactions (2002)

Zhang, B., Bao, Y. *„Cross-Dataset Learning for Age Estimation”* IEEE Access (2022)

Źródła Danych

Zdjęcia twarzy wykorzystane w modelu – UTK Faces <https://susanqq.github.io/UTKFace/>

Spis tabel i wykresów

Załączniki

Załącznik 1. Kod źródłowy

Streszczenie

Praca koncentruje się na automatycznej ocenie wieku na podstawie obrazów twarzy za pomocą konwolucyjnych sieci neuronowych (CNN). Rozwój technologii głębokiego uczenia umożliwia coraz dokładniejsze analizy biometryczne, co ma szerokie zastosowanie w takich dziedzinach jak handel detaliczny, medycyna czy bezpieczeństwo publiczne. Praca przedstawia przegląd literatury oraz dostępnych zbiorów danych, które mogą być wykorzystane do budowy modeli estymujących wiek. W ramach pracy zaproponowano stworzenie modelu głębokiej sieci neuronowej, który na podstawie zdjęcia twarzy określi, czy dana osoba jest pełnoletnia. Proces modelowania obejmuje przygotowanie i przetworzenie danych, wybór odpowiedniej architektury CNN oraz ewaluację skuteczności modelu. Praca omawia również potencjalne zastosowania stworzonego systemu w praktycznych kontekstach, takich jak implementacja wytrenowanego modelu w postaci systemu ułatwiającego weryfikację wieku użytkowników.