Politechnika Śląska
Wydział Automatyki, Elektroniki i Informatyki

Rozprawa doktorska

mgr inż. Magdalena Pawlyta

Klasyfikacja zachowań postaci ludzkiej z wykorzystaniem uczenia maszynowego na podstawie trajektorii punktów charakterystycznych w reprezentacji 3D i 2D

Promotor: prof. dr hab. inż. Konrad Wojciechowski,
Promotor pomocniczy: dr inż. Przemysław Skurowski,

Gliwice 2024

Spis treści

1 Wstęp

Analiza i klasyfikacja zachowań człowieka obserwowanych w strumieniu wideo pochodzącym z pojedynczej kamery jest problemem istotnym dla zastosowań praktycznych i równocześnie trudnym algorytmicznie. Trudność ta wynika z faktu, iż dane dostępne w strumieniu wideo powstają w wyniku rzutowania perspektywicznego sceny 3D, w tym ruchu człowieka, na płaszczyznę płytki CCD kamery. Parametry takiego rzutowania są nieznane oraz zmieniają się w trakcie ruchu – zależą one od orientacji człowieka, jako trójwymiarowej bryły, względem kamery. Rejestrowana osoba może znajdować się w różnej odległości mierzonej od ogniska kamery jak również znajdować się prostopadle do osi kamery lub być zwrócona do niej bokiem. Ponadto, w trakcie ruchu, odległość oraz orientacja względem osi kamery ulegają zmianie. W konsekwencji w pojedynczej ramce strumienia znajduje się sylwetkę człowieka, która może zmieniać wysokość i szerokość w wyniku ruchu.

W literaturze istnieją dwa główne podejścia do rozwiązywania przedstawionego problemu. Pierwsze zakłada odtworzenie modelu 3D człowieka poprzez dopasowanie modelu 3D do sylwetek obserwowanych w kolejnych ramkach wideo, co prowadzi do wielowymiarowego procesu minimalizacji. Drugie zakłada generowanie trajektorii punktów charakterystycznych sylwetki reprezentujących te same punkty modelu 3D, co nie zawsze jest możliwe do osiągnięcia. Zbiór tych trajektorii stanowi podstawę do klasyfikacji ruchu czy zachowania. Jednakże to podejście jest bardzo wrażliwe na: poprawność odnajdywania tego samego punktu modelu 3D w kolejnych ramkach wideo; znikanie śledzonego punktu; zależność kształtu i długości trajektorii od położenia; orientacji względem kamery.

Dodatkowo w celu budowy i oceny klasyfikatora rozpoznającego poszczególne zachowania jednej bądź większej liczby osób, konieczne jest uzyskanie jak największej liczby nagrań zawierającej te same ruchy rejestrowane przy różnych ustawieniach i pozycjach kamery. Zadanie to jest niezwykle trudne do zrealizowania – wymaga ogromnej liczby zsynchronizowanych ze sobą kamer wideo. Jednakże niezależnie od liczby wykorzystanych kamer niemożliwym jest uzyskanie nagrania z tej samej pozycji kamery dla różnych jej parametrów.

W celu uzyskania widoku z dowolnej perspektywy, z dowolnymi ustawieniami kamery wykorzystano technologię Motion Capture. Umożliwia ona rejestrację trójwymiarowych trajektorii wybranych punktów reprezentowanych przez markery, umieszczone w poszczególnych punktach anatomicznych (np. na stawach). Następnie za pomocą modelu wirtualnej kamery można dokonać rzutowania perspektywicznego, uzyskując tym samym idealne trajektorie punktów w przestrzeni dwuwymiarowej. Wirtualna kamera ta może być umieszczona w wybranej odległości i orientacji względem osoby.

Dodatkową zaletą wykorzystania systemów Motion Capture jest stosunkowa duża liczba śledzonych punktów. W zależności od wykorzystywanego modelu omarkerowania, śledzonych jest jednocześnie do 53 punktów dla jednej osoby. Uzyskanie takiej liczby trajektorii przy użyciu kamer wizyjnych jest niezwykle trudne, a w pewnych przypadkach wręcz niemożliwe. Konieczne jest, zatem przeprowadzenie redukcji znaczników w danych wejściowych w celu ustalenia ich optymalnej liczby.

Aspekt dotyczący rozpoznawania poszczególnych zachowań postaci ludzkiej można podzielić na dwa osobne zagadnienia. Rozpoznawanie samych akcji, oraz

rozpoznawanie cech osób wykonujących daną akcję. W medycynie od wielu lat wykorzystuje się analizę ruchu do oceny stanu zdrowia pacjenta. Najczęściej spotykana jest ocena chodu, obejmujące ocenę pracy poszczególnych stawów i mięśni pacjenta. Istnieje szereg wzorców prawidłowej pracy poszczególnych stawów podczas chodu. Wśród pacjentów o podobnym stanie fizycznym, można zauważyć charakterystyczny sposób odbiegania od tych wzorców – przykładowo u osób po endoprotezoplastyce stawu biodrowego praktycznie nie występuje ruch przywiedzenia operowanego stawu.

1.1 Zakres prac

Badanie przeprowadzone na potrzeby niniejszej pracy obejmowały szereg akwizycji ruchu osób z naniesionymi markerami, realizujących zachowania takie jak: chód, bieg, schylanie się, uderzenie czy kopanie. Dane te następnie zostały oczyszczone ze wszelkich nieprawidłowości, oraz poddane dodatkowej normalizacji. Pozyskane dla poszczególnych zachowań trajektorie markerów utworzą bazę danych 3D, która następnie posłuży, jako źródło bazy danych 2D. Do generowania dwuwymiarowych danych wykorzystane zostanie rzutowane perspektywiczne z parametrami zewnętrznymi i wewnętrznymi kamery. Zestaw trajektorii odpowiadających ustalonym parametrom będzie podstawą dla zarówno uczenia jak i rozpoznawania klasyfikatora. Zaletą planowanego podejścia badawczego jest możliwość taniego wytworzenia dowolnie licznych zbiorów, co ma znaczenie w przypadku klasyfikatorów typu DNN. Przeprowadzone badania pozwolą ustalić optymalną liczbę trajektorii dla poszczególnych rodzajów klasyfikatorów zarówno w przestrzeni trójwymiarowej jak i dwuwymiarowej oraz wykażą, jaki jest wpływ rzutowania perspektywicznego, na jakość klasyfikacji.

1.2 Układ pracy

Niniejsza praca została podzielona na siedem rozdziałów. Pierwszy rozdział stanowi wstęp, w którym opisano motywacje oraz sformułowano cele rozprawy. Drugi rozdział zawiera przegląd literatury wraz z szczegółowym opisem sposobu reprezentacji danych oraz przeglądem najpopularniejszych, publicznie dostępnych baz danych ruchu człowieka. W rozdziale trzecim przedstawiono metody głębokiego uczenia, oraz opisano dwa najpopularniejsze modele głębokich sieci neuronowych. Rozdział czwarty zawiera szczegółowy opis procesu tworzenia bazy danych, wykorzystanej w niniejszej pracy. Opis ten rozpoczyna się od wyjaśnienia, czym jest technologia Motion Capture. Następnie przedstawione zostają wybrane akcje proste, oraz metody wykorzystane do standaryzacji danych. Na końcu tego rozdziału przedstawiono metodę generowania dancyh dwuwymiarowych. Rozdział piąty skupia się na klasyfikacji wybranych akcji prostych w przestrzeni trójwymiarowej z wykorzystaniem dwóch różnych modeli głębokich sieci neuronowych. Zbadany zostaje wpływ rozmiaru wektora wejściowego, na jakość klasyfikacji (określaną za pomocą kilku różnych miar), wraz z porównaniem oby wybranych modeli. Rozdział szósty skupia się na określeniu granicy rozpoznawalności poszczególnych akcji prostych w przestrzeni dwuwymiarowej. Rozdział ten skupia się na analizie wpływu rzutowania perspektywicznego na rozpoznawalność danej akcji prostej, oraz błędy w klasyfikacji. Ostatni rozdział jest podsumowaniem, w którym opisana jest realizacja celów badawczych i naukowych. Przedstawione zostają również kierunki możliwych dalszych prac.

2 Problem klasyfikacji zachowań człowieka

Klasyfikacja zachowań postaci ludzkiej jest dynamicznie rozwijanym, bardzo złożonym problemem badawczym, znajdującym zastosowanie w wielu dziedzinach takich jak system monitoringu i nadzoru [1–4], opieka zdrowotna [5–7], interakcja człowiek-robot [8–10], wspomaganie treningów sportowych [11–13], czy zastosowania wojskowe [14].

Problem klasyfikacji zachowań postaci ludzkiej można podzielić na kilka pomniejszych problemów, z których każdy może stanowić oddzielny obszar badawczy. Są to kolejno: rejestracja danych, ekstrakcja cech, segmentacja, oraz sama klasyfikacja. Na przestrzeni ostatnich lat powstało wiele różnych artykułów podsumowujących oraz porównujących wybrane aspekty każdego z wymienionych problemów [1, 3, 15–23].

Zachowanie człowieka jest bardzo szerokim pojęciem. Może dotyczyć zarówno wykonywania prostych gestów, jak i złożonych aktywności. W literaturze można spotkać kilka różnych podziałów zachowań człowieka w zależności od stopnia ich złożoności.

Aggarwal i Ryoo [1] zaproponowali podział zachowań człowieka na cztery podgrupy. Pierwszą z nich stanowią proste gesty. W czasie ich wykonywania zmianie ulega położenie pojedynczego stawu lub kończyny. Przykładem gestu może być wymach ręką. Drugą grupę stanowią akcje, które składają się z kilku podstawowych ruchów. Przykładem takiej akcji może być chód. Ostatnie dwie grupy stanowią złożone zachowania obejmujące interakcje kolejno pomiędzy człowiekiem a obiektem (np. podnoszenie obiektu z podłoża), oraz pomiędzy dwoma lub większą liczbą ludzi (np. rozmowa).

Moeslund i inni [24] zaproponowali podział na trzy grupy zachowań. Pierwsze dwie grupy są identyczne jak w poprzednim przykładzie - gesty i proste akcje. Trzecią grupę stanowią natomiast zachowania będące połączeniem kilku akcji, przykładem takiego zachowania może być bieg połączony ze skokiem.

W niniejszej pracy rozpoznawane zachowania nie są złożone, co odpowiada definicji Akcji wprowadzonych we wcześniej wspomnianych artykułach.

2.1 Sposoby reprezentacji danych

Odpowiednia selekcja cech oraz ich reprezentacja jest klasycznym problemem uczenia maszynowego [25]. W przypadku rozpoznawania zachowań postaci ludzkiej w strumieniu wideo należy nie tylko stwierdzić czy i gdzie w obrazie znajduje się osoba, ale również określić zmiany jej położenia w czasie i przestrzeni. W literaturze można wyróżnić kilka różnych podejść do rozwiązania tego problemu.

Jedno z najpopularniejszych podejść bazuje na estymacji szkieletowej, oraz detekcji poszczególnych stawów. Estymacja ta odbywa się najczęściej na dwa różne sposoby. Pierwszy z nich przechodzi od ogółu do szczegółu - najpierw dokonywana jest detekcja postaci ludzkiej w obrazie, a następie estymowane są poszczególne części ciała [26–28]. Drugi działa odwrotnie i zaczyna od detekcji poszczególnych części ciała, a następnie łączy je w całość [27, 29, 30]. Kolejne kroki w tym podejściu obejmują kolejno estymację postaci, ekstrakcję cech a następnie klasyfikację.

Drugie bardzo popularne podejście, bazuje na śledzeniu poszczególnych punktów charakterystycznych w kolejnych ramkach obrazu [31–35]. Utworzony w ten

sposób zbiór trajektorii stanowi źródło dla klasyfikatora. Do śledzenia wykorzystuje się między innymi przepływy optyczne (ang. optical flow), albo filtry Kalmana. Przy czym podejście to jest bardzo wrażliwe na zmiany otoczenia, czy okluzje. W niektórych przypadkach wymagane jest zastosowania dodatkowych algorytmów, w celu korekty uzyskanych trajektorii. W przypadku tego podejścia kolejne kroki obejmują kolejno detekcję obiektu, jego śledzenie i klasyfikację.

Tak samo jak w większości obszarów badawczych, tak również w tym do ekstrakcji cech wykorzystuje się również metody głębokiego uczenia. Najczęściej spotykane podejście zakłada wykorzystanie konwolucyjnych lub rekurencyjnych sieci neuronowych na surowych danych wideo w celu automatycznej ekstrakcji cech, a następnie klasyfikacji konkretnych nagrań.

Sama klasyfikacja, niezależnie od sposobu estymacji cech, może się obywać za pomocą różnego rodzaju klasyfikatorów takich jak: wektory podpierające (ang. Support Vector Machines, SVM), lasy losowe (ang. Random Forests), czy sieci neuronowe zarówno tradycyjne jak i głębokie.

Ponieważ celem niniejszej pracy nie jest, jak w przypadku wielu dostępnych publikacji, znalezienie najlepszej metody klasyfikacji konkretnych akcji, a zbadanie wpływu rzutowania perspektywicznego na samą, jakość klasyfikacji, zdecydowano ograniczyć liczbę klasyfikatorów do dwóch wariantów dość generycznych głębokich sieci neuronowych, które zostały opisanych w rozdziale 3.

2.2 Dostępne bazy danych

Niezależnie od sposobu reprezentacji danych, czy wyboru metody klasyfikacji, najważniejszą rzeczą jest sam dostęp do danych ruchu. Tak samo jak w przypadku większości problemów klasyfikacyjnych istnieje szereg publicznych baz danych ruchu. Szczegółowe podsumowanie baz danych ruchu, dostępnych w danym przedziale czasowym przedstawiono kolejno w [36] i [37]. Publicznie dostępne bazy danych można podzielić na dwie kategorie w zależności od typu nagrań - kontrolowane, wykonywane w ściśle określonych warunkach oraz niekontrolowane, będące fragmentami nagrań z filmów bądź nagrań z kamer monitoringu.

Do tej pierwszej kategorii należą dwie bardzo popularne bazy danych KHT [38] i Weizmann [39]. Baza Weizmann zawiera 9 różnych akcji (chód, bieg, podskoki na dwóch nogach, pajacyki, skok w przód, wymachy jedną/dwoma rękami, klęknięcie oraz krok odstawno-dostawny), wykonanych przez 9 różnych osób. Baza KHT zawiera tylko 6 akcji (chód, trucht, bieg, uderzenie, machanie ręką, klaskanie), wykonanych przez 25 różnych osób. Łącznie w bazie tej znajduje się 2391 nagrań. Dodatkowo nagrania te zostały wykonane w 4 różnych wariantach - na zewnątrz, na zewnątrz w innych ubraniach, na zewnątrz ze zmianą położenia względem kamery oraz w pomieszczeniu.

Znacznie liczniejsza grupę stanowią bazy danych, w których konkretne nagrania są częścią filmów kinowych, programów telewizyjnych czy filmów z YouTube. Należą do nich między innymi kolejne wariacje baz UFC, począwszy od UFC11 [40], UCF50 [41] i UCF101 [42]. Numer przy nazwie danej bazy odpowiada liczbie różnych akcji w niej zawartych. Wraz z rosnącą liczbą różnych akcji rósł też poziom ich szczegółowości. Baza UFC11 zawiera akcje stosunkowo proste jak chód, czy bieg, podczas gdy UFC101 zawiera bardzo specyficzne akcje jak nakładanie makijażu, obcinanie włosów, gra w krykieta czy raczkowanie. Nagrania dostępne w tych bazach zostały przede wszystkim wyodrębnione

z różnych filmów z platformy YouTube. Podobne źródło danych zostało wykorzystane przy tworzeniu kolejnych wariantów bazy Kinetics - Kinetics 400 [43], Kinetics 600 [44] oraz Kinetics 700 [45], zawierające coraz większą liczbę klas (ich liczba określona jest w nazwie bazy).

Inną bardzo popularną bazą danych, której źródło stanowiły głównie filmy z platformy YouTube jest baza HMDB51 [46]. Zawiera ona 51 różnych rodzajów akcji takich jak chód, bieg, machanie, uderzenia, ale również bardziej złożone i nietypowe jak walka na miecze.

Baza danych Hollywood2 [47] powstała natomiast na podstawie 69 różnych filmów kinowych. Zawiera 12 różnych codziennych akcji takich jak jedzenie, bieganie, chodzenie wraz z informacją o miejscu dziania się akcji (biegnie po drodze/chodniku, jedzenie w kuchni/w kawiarni). Podobnie jak baza MultiTHUMOS [48] zawierająca 30 godzin nagrań pochodzących w różnych filmów, z podobnymi akcjami codziennymi.

Baza Olympic Sports [49] natomiast skupia się na rozpoznawaniu 16 różnych akcji sportowych takich jak różnego rodzaju skoki (wysoki, w dal), rzut młotem, czy gra w tenisa. Źródłem nagrań są filmy z Olimpiady dostępne na YouTube.

W grupie tej możemy też wyróżnić dodatkową podgrupę, w której źródło danych stanowią miejskie kamery bezpieczeństwa. Należą do nich między innymi baza RWF-2000 [50] zawierająca akcje niebezpieczne. UCF-Crime [51] zawierająca 128 godzin filmów z kamer wizyjnych, podzielonych na 1900 osobnych nagrań zawierających 120 różnych akcji niebezpiecznych takie jak bójki, napady, rabunki oraz akcje neutralne takie jak chód czy bieg. Największą bazą danych w tej kategorii jest baza VMASS2 [52] zawierająca 4000 godzin nagrań z kamer monitoringu wraz z opisem. Zawiera 400 różnych akcji takich jak chodzenie, rozmowa, jazda na rowerze, bieg, zabawa, upadek, wychodzenie/wchodzenie do budynku, palenie papierosów, rozmowa telefoniczna, czy uciekanie.

Istnieje znacznie więcej publicznie dostępnych baz danych - część badaczy na potrzeby konkretnych eksperymentów tworzy swoje własne, małe bazy danych zawierające kilka konkretnych ruchów wykonanych przez niewielką liczbę osób. Podsumowanie wszystkich omówionych wcześniej baz znajduje się w tabeli 1.

Tabela 1: Najpopularniejsze, publicznie dostępne bazy danych

Nazwa bazy danych	Liczba		
	nagrań	aktorów	akcji
KHT [38]	2391	25	6
Weizmann [39]	81	9	9
Hollywood2 [47]	1707	-	12
Olympic Sports [49]	800	-	16
HMDB51 [46]	7000	-	51
UCF101 [42]	>13.000	-	101
VMASS2 [52]	>6.000.000	-	400
UCF-Crime [51]	1900	-	13
MultiTHUMOS [48]	400	-	65
NTU RGB+D 120 [53]	>114.000	-	120
Kinetics 700 [45]	650.000	-	700
RWF-2000 [50]	2000	-	2

Większość wymienionych baz danych zawiera kilka podobnych akcji codziennych (chód, czy bieg). Dodatkowo w przypadku niekontrolowanych nagrań akcje te są widoczne z różnych perspektyw. Konkretna akcja może być lepiej lub gorzej klasyfikowana w zależności od orientacji danej osoby względem kamery. Co więcej rodzaj pomyłek również może być inny w zależności od położenia i parametrów kamery. W celu zbadania wpływu rzutowania perspektywicznego nie tylko na ogólną, jakość klasyfikacji, ale też na rodzaje błędów, do jakich może dochodzić niezbędne jest stworzenie bazy zawierającej ten sam ruch widziany z wielu różnych perspektyw. Uzyskanie takiej bazy danych przy wykorzystaniu kamer wizyjnych jest praktycznie niemożliwe. Dlatego też na potrzeby zaplanowanych eksperymentów utworzono nową bazę danych, wykorzystując w tym celu technologię Motion Capture (szczegółowy opis w rozdziale 4.1). Pozwoli to nie tylko na uzyskanie dowolnej liczby perspektyw, z których dany konkretny ruch będzie widoczny, ale również zagwarantuje idealne niczym niezakłócone dane. Dzięki temu badany będzie wpływ rzutowania perspektywicznego na sam rodzaj akcji oraz sposób jej wykonania, bez wpływu błędów pochodzących z metod estymacji szkieletowej, czy śledzenia punktów charakterystycznych, które czasem wymagają dodatkowych algorytmów uzupełniających braki w danych.

3 Metody głębokiego uczenia

3.1 Głębokie uczenie

Głębokie uczenie jest jedną z metod uczenia maszynowego opartą na sieciach neuronowych. Samo słowo głębokie odnosi się do dużej liczby warstw w sieci. Tak jak tradycyjne sieci neuronowe, głębokie sieci można uczyć metodą nadzorowaną, częściowo nadzorowaną, lub całkowicie bez nadzoru [54]. Głębokie uczenie, ze względu na swoje rezultaty znajduje zastosowanie w coraz większej liczbie dziedzin, od przetwarzania obrazów, poprzez rozpoznawanie mowy, tłumaczenia, bioinformatykę, medyczną analizę obrazów, czy nawet tworzenie nowych obrazów na podstawie już istniejących zdjęć, albo szczegółowego opisu.

Metody głębokiego uczenia w wielu aspektach przewyższają i tym samym wypierają dotychczasowe rozwiązania. Dzieje się tak, dlatego, że metody te wykorzystują bardzo złożone sieci neuronowe, których zadaniem jest nie tylko klasyfikacja, ale również ekstrakcja cech z wektora wejściowego. Umożliwia to uzyskanie znacznie większej ilości, niekiedy dość abstrakcyjnych, cech, których nie byłby w stanie wskazać ludzki ekspert. Metody pozyskiwania cech różnią się w zależności od zastosowanej architektury sieci i zostały bardziej szczegółowo opisane dla dwóch wybranych architektur w podrozdziałach 3.2 i 3.3.

3.1.1 Rys historyczny

Historia głębokiego uczenia jest powiązana z historią uczenia maszynowego i sięga do lat 1943, gdy Walter Pitts i Warren McCulloch stworzyli pierwszy model wzorowany na pracy ludzkiego mózgu [55]. Moment ten można uznać za początek rozwoju głębokiego uczenia. Kolejnym przełomowym krokiem było opracowanie przez Franka Rosenblatt'a modelu perceptronu [56] - był to algorytm rozpoznawania wzorców oparty na dwuwarstwowej sieci komputerowej wykorzystujący proste dodawanie i odejmowanie. W 1980 Kunihiko Fukushima zaproponował kolejny model - neocognitron [57]. Była to wielowarstwowa, hierarchiczna sieć neuronowa używana między innymi do rozpoznawania pisma ręcznego. Sieć ta stała się w późniejszych latach inspiracją dla stworzenia splotowych sieci neuronowych (ang. Convolutional Neural Network, CNN). Z kolei pierwsze rekurencyjne sieci neuronowe (ang. Recurrent Neural Networks RNNs) zaproponowane zostały w 1986 roku przez Jordana [58]. Kolejnym ważnym krokiem w rozwoju uczenia głębokiego, było zaproponowanie w 1998 roku sieci LeNet [59]. Jednakże ze względu na znaczne ograniczenia, w szczególności sprzętowe, sieć ta była dość prosta i nie można jej było zastosować na dużych zbiorach danych, przez co nie zyskała ona dużego uznania.

Największy przełom dla głębokiego uczenia nastąpił po roku 2006. Przyczyniła się do tego między innymi zaproponowana przez Hintona i innych [60] sieć "głębokich przekonań" (ang. Deep Belief Networks DBN), wraz z algorytmem uczenia wstępnego. Algorytm ten zakładał wstępne szkolenie sieci, po którym następowało zamrożenie jej parametrów. Następnie dokładana była kolejna warstwa sieci i w następnym cyklu uczenia, dobierane były wartości parametrów tylko tej nowej warstwy. Umożliwiło to szkolenie znacznie głębszych i bardziej rozbudowanych sieci.

Wraz ze wzrostem głębokości sieci, pojawił się kolejny problem. W trackie procesu uczenia, niższe warstwy nie były uczone przez warstwy wyższe, gdyż

nie docierał do nich żaden sygnał uczenia. Problem ten dotyczył sieci, których uczenie opierało się na gradiencie i został nazwany problemem zanikającego gradientu. Problem ten próbowano rozwiązać za pomocą wielu metod, takich jak walidacja grupy przykładów (ang. Batch normalization), czy zmiana funkcji aktywacji neuronu na funkcje ReLU [61]. Sposoby uczenia sieci neuronowych zostały nieco szerzej opisane w podsekcji 3.1.2.

W następnych latach znikały kolejne ograniczenia takie jak czas trwania obliczeń, czy dostęp do odpowiednio dużych baz danych. Do eliminacji pierwszego z wymienionych aspektów przyczynił się rozwój technologiczny, a w szczególności rozwój procesorów graficznych. Ich wykorzystanie w znaczący sposób przyspieszało proces uczenia. Dodatkowo na przestrzeni lat opracowano różne algorytmy rozproszonego uczenia, co jeszcze bardziej skracało czas uczenia się sieci [62,63]. Eliminacją drugiego aspektu zajęli się naukowcy z całego świata, opracowując ogromne, publicznie dostępne bazy danych. Przykładem takiej bazy, może być baza ImageNet opublikowana w 2009 roku przez Fei-Fei Li [1]. Baza ta aktualnie zawiera ponad 14 milionów zdjęć w wysokiej rozdzielczości. Na jej podstawie stworzono wiele popularnych, wstępnie przetrenowanych sieci takich jak AlexNet [64] czy ResNet [65].

3.1.2 Proces uczenia sieci neuronowej

Proces uczenia się sieci neuronowych sprowadza się do modyfikacji współczynników wagowych połączeń pomiędzy poszczególnymi neuronami, tak, aby zminimalizować błąd predykcji. Wyróżnia się trzy metody uczenia się sieci:

- uczenie nadzorowane - sieci prezentowane są dane wejściowe, oraz odpowiadające im sygnały wyjściowe (np. obrazek zwierzęcia, oraz informacja, co to za zwierzę).

- uczenie nienadzorowane - sieci prezentowane są tylko dane wejściowe, a jej zadaniem jest stworzenie własnych kategorii/klas.

- uczenie częściowo nadzorowane - połączenie dwóch poprzednich metod, sieci prezentowane są zarówno pary danych wejście-wyjście, jak i nieoznaczone.

Niezależnie od wybranej metody, potrzebny jest algorytm, który na podstawie błędów popełnianych przez sieć będzie modyfikował wagi połączeń, tak, aby minimalizować błędy. Pierwszą wersje takiego algorytmu zaproponował już w 1960 roku Henry J. Kellego [66]. Algorytm ten, nazywany algorytmem wstecznej propagacji błędu (ang. Backpropagation algorithm), został potem udoskonalony i spopularyzowany przez Davida E. Rumelharta, Geoffreya E. Hintona, oraz Ronalda J. Williamsa w 1986 roku [67].

Za pomocą algorytmu oblicza się gradient funkcji straty z uwzględnieniem poszczególnych wag połączeń sieci dla pojedynczej pary wejście-wyjście. Gradient obliczany jest w jednej warstwie na raz, w kolejnych krokach przechodzą wstecz aż do ostatniej warstwy. Następnym etapem jest znalezienie minimum funkcji gradientu. Jedną z popularnych metod jest metoda stochastycznego spadku wzdłuż gradientu [68].

[1] https://www.image-net.org/

W przypadku głębokich sieci neuronowych najczęściej wykorzystuje się rozszerzenie stochastycznego spadku - optymalizator Adam (ang. Adaptive moment estimation) [69]. Autorzy połączyli w nim zalety dwóch innych rozszerzeń stochastycznego spadku wzdłuż gradientu: algorytmu adaptacyjnego gradientu (ang. Adaptive Gradient Algorithm, AdaGrad) [70] oraz propagację średniej kwadratowej (ang. Root Mean Square Propagation RMSProp) [71].

Innym ważnym aspektem uczenia sieci neuronowych jest określenie tak zwanej funkcji błędu. Jej zadaniem jest zmierzenie różnicy pomiędzy predykcją sieci a prawidłowymi wynikami. Wyróżnić można dwie najczęściej stosowane funkcje błędu, wykorzystywane do rozwiązywania różnych problemów. Pierwszą z nich jest błąd średniokwadratowy (ang. Mean Squared Error, MSE), który oblicza średnią sumę błędów pomiędzy wektorem wyjściowym (y) a wynikami prawidłowymi ($\hat{y}$):

$$MSE = \frac{1}{N}\sum_{i=0}^{N}(y_i - \hat{y}_i)^2 \tag{1}$$

Funkcja ta jest najczęściej wykorzystywana przy problemach regresyjnych. W przypadku zadań klasyfikacyjnych najczęściej wykorzystuje się entropię krzyżową (ang. Cross-Entropy, CE). Jej zadaniem jest ocena nietrafności rozkładu danych wyjściowych względem rozkładu danych prawdziwych. Matematycznie można ją wyrazić za pomocą następującego wzoru:

$$CE = -\sum_{i=0}^{N}\hat{y}_i \cdot log(y_i) \tag{2}$$

Funkcja ta została wykorzystana w niniejszej pracy.

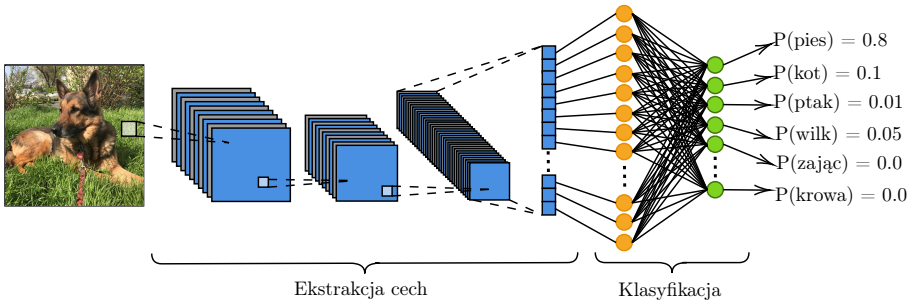
3.2 Konwolucyjne sieci neuronowe

Konwolucyjne sieci neuronowe, to rodzaj sieci zaprojektowanych głównie do przetwarzania obrazów. Jak sama nazwa wskazuje, wykorzystywana jest w nich operacja konwolucji, w co najmniej jednej warstwie. Sama idea tego rodzaju architektury nie jest nowa - po raz pierwszy została zaproponowana przez K. Fukushime w 1988 roku [72]. Ze względu na spore ograniczenia sprzętowe, które utrudniały jej uczenie, sieć ta nie zyskała zbytniej popularności. Kilka lat później uległo to zmianie, gdy w 1998 roku Y. LeCun i inni wytrenowali ten rodzaj architektury, wykorzystując algorytmu bazujący na gradiencie, do rozpoznawania problemu rozpoznawania pisanych odręcznie cyfr [59]. Przyczyniło się to do znaczącej popularyzacji tego rodzaju sieci. W kolejnych latach badacze z różnych dziedzin dokonywali dalszych udoskonaleń tego modelu, osiągając coraz lepsze rezultaty, zwłaszcza w przypadku przetwarzania i klasyfikacji obrazów.

3.2.1 Zasada działania sieci konwolucyjnych

Sieć CNN można podzielić na dwie główne części: ekstraktor cech, oraz klasyfikator (rys. 1). Zadaniem ekstraktora jest zastąpienie ludzkiego operatora i znalezienie w danych wejściowych różnego rodzaju cech, które potem stanowią dane wejściowe dla klasyfikatora. Wśród warstw składających się na ekstraktor wyróżnia się dwie podstawowe warstwy następujące bezpośrednio po sobie: konwolucyjną (splotową) oraz próbkującą (ang. pooling). Wyjścia z tych warstw

tworzą dwuwymiarową tablice wartości zwaną mapą cech. Jako klasyfikator najczęściej wykorzystywana jest tradycyjna, w pełni połączona sieć neuronową ze sprzężeniem zwrotnym (ang. feed-forward neural network, FNN).



Rysunek 1: Schemat konwolucyjnej sieci neuronowej

W warstwach konwolucyjnych zachodzi operacja splotu pomiędzy mapą cech z poprzedniej warstwy a filtrami (jądrami). Filtry przesuwane są wzdłuż danych wejściowych w pionie oraz poziomie. Operacje splotu można zapisać jako:

$$s(i,j) = \sum_x \sum_y f(x,y) \cdot h(i-x, j-y) \tag{3}$$

gdzie x to sygnał wejściowy, a f to filtr.

Następnie wyjścia z każdego filtru przechodzą przez różnego rodzaju funkcje aktywacyjne takie jak funkcja sigmoidalna (4), czy ReLU (5) tworząc wspomnianą wcześniej mapę cech. Wartości poszczególnych filtrów dobierane są w procesie uczenia się sieci.

$$S(x) = \frac{1}{1+e^{-x}} \tag{4}$$

$$R(x) = max(0,x) \tag{5}$$

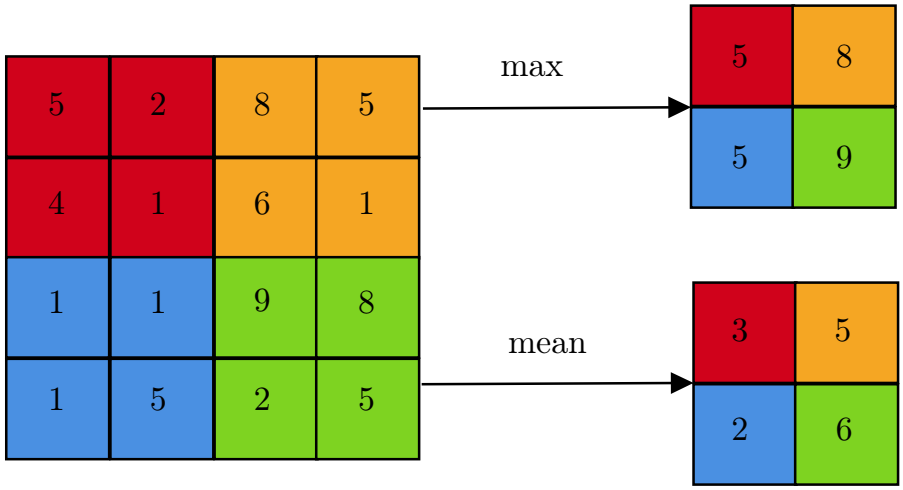
Zadaniem warstwy próbkującej jest zmniejszenie rozmiaru map wejściowych. W warstwie tej liczba map pozostaje stała. Ponownie stosowane jest jądro (maska) o rozmiarach N x N określające o ile ma zostać zmniejszony rozmiar map wejściowych. Zmniejszenia dokonuje się najczęściej za pomocą jednej z dwóch operacji - uśredniania, lub maksymalizacji. Wartości z jądra zostają kolejno uśrednione lub wybierana jest spośród nich jedna największa wartość, która stanowi odpowiednik mapy wejściowej na mapie wyjściowej. Operacje te można zapisać za pomocą następujących wzorów:

$$P_{j,m} = mean(h_{j,(m-1)N+r}) \tag{6}$$

$$P_{j,m} = max(h_{j,(m-1)N+r}) \tag{7}$$

gdzie $N \in \{1,...,R\}$ to przesunięcie pomiędzy kolejnymi próbkowanymi obszarami, gdzie $N < R$.

Od rozmiaru użytej maski zależy stopnień zmniejszenia się mapy. Przykładowo zastosowanie maski o rozmiarach *2 x 2* zmniejszy mapę wejściową dwukrotnie (rys. 2). Zmniejszone mapy stanowią następnie wejście dla kolejnych warstw splotowych.



Rysunek 2: Wizualizacja operacji uśrednienia i maksymalizacji, dla jądra o rozmiarze 2x2, z krokiem 2 w obu osiach

Zastosowanie wielu następujących po sobie warstw splotowych i próbkujących umożliwia odnajdywanie coraz bardziej złożonych wzorców - jest to zasadniczo analiza wielorozdzielcza. Warstwy niższego poziomu odpowiadają za ekstrakcje prostych cech, w przypadku obrazu może to być rozpoznawanie pionowych czy poziomych linii. Kolejne warstwy łączą cechy znalezione przez warstwy niższego poziomu, w bardziej złożone jak konkretne kształty począwszy od prostych figur geometrycznych, skończywszy na twarzach czy samochodach w najwyższych warstwach.

3.2.2 Sekwencyjne dane wejściowe

Sieci konwolucyjne próbowano, z większym lub mniejszym sukcesem, zaadaptować do problemów w których dane wejściowe stanowią nie obrazy a jednowymiarowe sekwencje, np. dane z różnych czujników. Wśród wielu sposobów rozwiązania tego problemu można wyróżnić dwa główne podejścia.

Pierwsze zakłada konwersję sygnału 1D na 2D. Przykładem takiej konwersji może być metoda zaproponowana przez W. Jiang i innych, do rozpoznawania akcji postaci ludzkiej na podstawie danych z czujników [73]. W metodzie tej sygnały z czujników konwertowane były na poszczególne "wiersze" obrazu, gdzie wartość sygnału odpowiadała wartości kolejnego piksela. Utworzone w ten sposób obrazy stanowiły wejście dla sieci CNN.

Drugie, znacznie prostsze podejście bazuje na drobnej modyfikacji klasycznej sieci CNN - dane wejściowe stanowią pojedyncze sygnały, a filtry stosowane na kolejnych warstwach splotowych są jednowymiarowe.

3.3 Rekurencyjne sieci neuronowe

Rekurencyjne sieci neuronowe (ang. Recurrent Neural Networks, RNN), to rodzaj sieci, który zawiera dodatkowe, cykliczne połączenia, które umożliwiają naukę dynamiki czasowej w danych sekwencyjnych. Implementację tej idei przedstawił w 1990 J. Elman [74]. W zaproponowanej przez niego architekturze, wyjścia z warstw ukrytych stanowiły również wejścia wraz z normalnymi danymi wejściowymi ukrytej warstwy, co matematycznie można wyrazić w następujący sposób:

$$h_t = \sigma_h(U_h x_t + W_h h_{t-1} + b_h) \tag{8}$$

$$y_t = \sigma_y(W_y h_t + b_y) \tag{9}$$

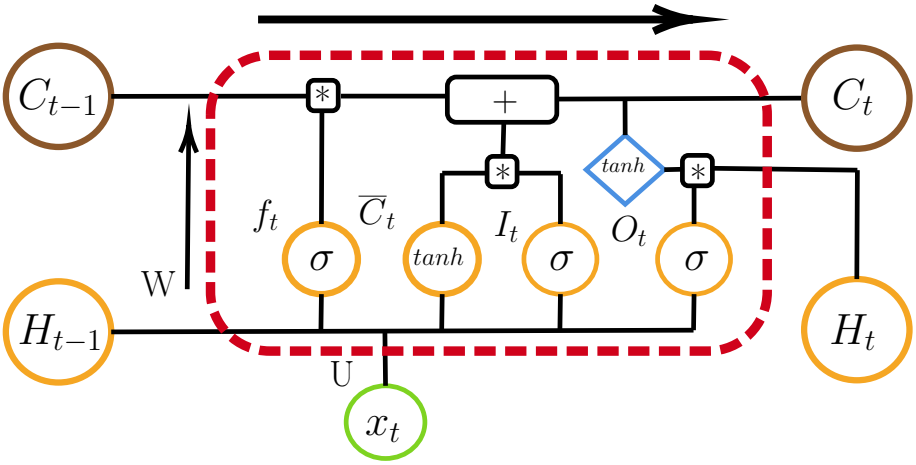
gdzie: h_t to aktualna wartość ukrytego stanu (pamięć), h_{t-1} to poprzednia wartość ukrytego stanu, x_t to aktualna wartość danych wejściowych, a U_h, W_h, W_y to wagi pomiędzy poszczególnymi połączeniami pomiędzy wejściami, a wyjściami warstw ukrytych.

Głównym problemem przy zastosowaniu w praktyce tej architektury jest wspomniany wcześniej problem zanikającego gradientu. Istnieje wiele rozwiązań tego problemu. W przypadku sieci RNN jednym z takich rozwiązań była modyfikacja sieci zaproponowana w 1997 roku przez S. Hochreiter i J. Schmidhuber. Wprowadzała ona specjalne komórki zwane LSTM (z ang. Long Short-Term Memory) [75]. Rozwiązanie to w późniejszych latach było wielokrotnie modyfikowane i udoskonalane.

3.3.1 Zasada działania sieci LSTM

Podstawą działania sieci LSTM, są specjalne komórki pamięci zawierające dodatkowe parametry oraz specjalne bramki rys. 3. Wyróżnia się następujące komponenty komórki:

- Bramka zapomnienia (f_t) - warunkowo decyduje, kiedy należy usunąć informacje z komórki.
- Bramka wejściowa (I_t) - kontroluje przepływ nowych informacji w komórce. Warunkowo decyduje, które informacje z wektora wejściowego zostaną wykorzystane do zaktualizowania pamięci komórki.
- Bramka wyjściowa (O_t) - na podstawie wejścia oraz aktualnej wartości pamięci komórki, warunkowo decyduje o wartości na wyjściu komórki.
- Stan ukryty (H_t) - zawiera informacje z poprzedniej próbki danych, odpowiada stanowi krótkoterminowemu.
- Wewnętrzny stan komórki (C_t) - odpowiada stanowi długoterminowemu.



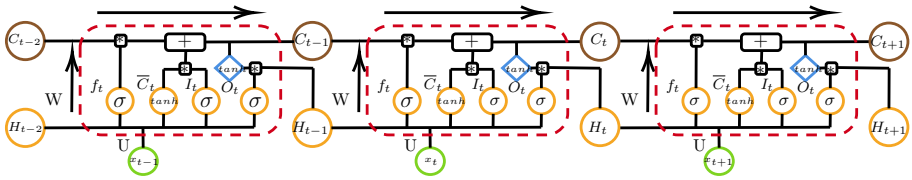
Rysunek 3: Schemat komórki LSTM

Matematycznie operacje na poszczególnych bramkach można zapisać w następujący sposób:

$$f_t = \sigma(x_t \cdot U_f + H_{t-1} \cdot W_f) \tag{10}$$
$$I_t = \sigma(x_t \cdot U_i + H_{t-1} \cdot W_i) \tag{11}$$
$$\overline{C}_t = tanh(x_t \cdot U_c + H_{t-1} \cdot W_c) \tag{12}$$
$$O_t = \sigma(x_t \cdot U_o + H_{t-1} \cdot W_o) \tag{13}$$
$$C_t = f_t \cdot C_{t-1} + \overline{C}_t \cdot I_t \tag{14}$$
$$H_t = tanh(C_t) \cdot O_t \tag{15}$$

gdzie: x_t to wektor wejściowy, H_{t-1} poprzednia wartość wyjścia z komórki, H_t aktualne wyjście z komórki, C_{t-1} poprzednia wartość pamięci komórki, C_t aktualna wartość pamięci komórki, W wektor wag dla bramek, U wektor wag dla danych wejściowych.

Sieć LSTM składa się z co najmniej jednej warstwy komórek (rys. 4). Na wyjściu z danej warstwy mogą się znajdować wyjścia z wszystkich komórek, co umożliwia mapowanie sekwencji na sekwencje. Do zadań klasyfikacji pod uwagę brana jest tylko wartość zwracana przez ostatnią komórkę. Podobnie jak w przypadku sieci CNN, po warstwach LSTM znajduje się klasyczna sieć FNN odpowiadająca za ostateczną klasyfikację danych.

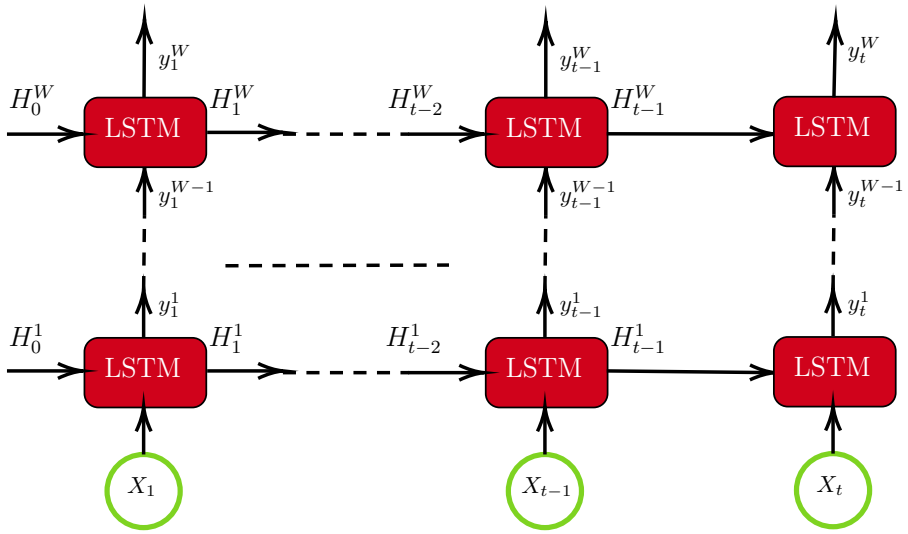


Rysunek 4: Fragment warstwy sieci LSTM

3.3.2 Wybrane architektury sieci LSTM

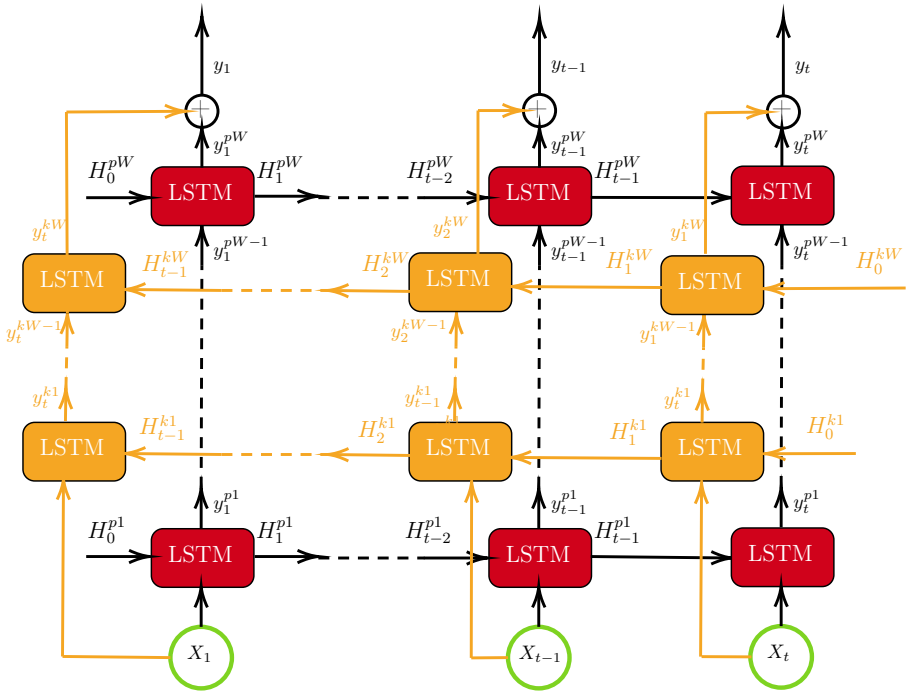
Istnieje wiele różnych architektur wykorzystujących komórki LSTM różniących się między sobą między innymi sposobami łączenia warstw, połączeniami między warstwami czy przepływem danych. Biorąc pod uwagę sposób przepływu danych wyróżnia się dwa podstawowe rodzaje sieci - jedno i dwukierunkową.

Przykładowy schemat jednokierunkowej sieci LSTM został przedstawiony na rysunku 5. Dane wejściowe dla pierwszej warstwy LSTM stanowi sekwencja próbek $(x_1, x_2, ..., x_t)$. Wyjścia z komórek niższego poziomu $(y_1^{W-1}, y_2^{W-1}, ..., y_t^{W-1})$ stanowią wejścia dla komórek w warstwach (W) wyższego poziomu. Początkowe wartości wszystkich ukrytych stanów (H_0), oraz wewnętrznych stanów komórki (C_0) inicjalizowane są za pomocą zera.



Rysunek 5: Schemat jednokierunkowej sieci LSTM

Model dwukierunkowy składa się z dwóch równoległych "podsieci" LSTM, o tej samej architekturze (liczba warstw, czy liczba komórek na warstwach). W pierwszej wektor danych odczytywany jest zgodnie z kolejnością (kolejne próbki czasowe, od lewej do prawej), w drugiej kolejność jest odwrócona. Na rysunku 6 przedstawiono przykładową sieć dwukierunkową. Ponieważ dane w obu podsieciach przepływają niezależnie od siebie, konieczne jest wprowadzenie dodatkowej operacji na wyjściu z sieci. W prezentowanym przykładzie na wyjściu z każdej komórki ostatniej warstw wartości z sieci działającej w przód i tył są sumowane.



Rysunek 6: Schemat dwukierunkowej sieci LSTM

3.4 Warstwy końcowe

Zarówno w przypadku sieci CNN jak i LSTM wykorzystywanych do zadań klasyfikacji, po warstwach w pełni połączonych następują kolejno tzw. warstwa softmax, oraz warstwa klasyfikująca. Zadaniem warstwy softmax jest przekształcenie surowych wyników z sieci neuronowej (x) w wektor prawdopodobieństwa przynależności do poszczególnych klas, co matematycznie można wyrazić w następujący sposób:

$$Softmax(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{N} e^{x_j}} \tag{16}$$

Umożliwia to określenie prawdopodobieństwa przynależności danych wejściowych do wszystkich zdefiniowanych klas. Następnie na podstawie tego prawdopodobieństwa w warstwie klasyfikującej obliczana jest funkcja błędu, oraz obliczane jest ostateczne wyjście z sieci.

3.5 Wybór architektur na potrzeby klasyfikacji zachowań człowieka

Zarówno sieci LSTM jak i CNN można modyfikować na wiele różnych sposobów. Począwszy od prostych zmian funkcji aktywacji, poprzez wprowadzanie dodatkowych połączeń, czy wręcz utworzenie sieci, która jest połączeniem obu wspomnianych sieci. Niektóre podejścia zakładają rezygnację z warstw w pełni

połączonych i zastąpienie ich innym rodzajem klasyfikatora. Możliwe jest również wykorzystanie jednej z dostępnych, wstępnie przetrenowanych sieci neuronowych takich jak ResNet [76], a następnie w procesie transferu wiedzy (ang. transfer learning) [77] dostosowanie jej do rozwiązania problemu klasyfikacji zachowań postaci ludzkiej.

Poza opisanymi modelami istnieją również inne, takie jak sieci GRU [78], które w swym działaniu przypominają sieci LSTM. Główna różnica pomiędzy tymi dwoma rodzajami sieci jest taka, że komórki GRU nie posiadają bramki wyjściowej, co zmniejsza liczbę parametrów uczenia sieci. Kolejnym, zyskującym coraz większą popularność, modelem jest sieć oparta o Transformery [79]. Początkowo sieci takie były wykorzystywane do zadań dotyczących przetwarzania języka naturalnego. Ich główną zaletą było to, iż usuwały sekwencyjną naturę problemu - sekwencja słów stanowiła wejście w całości, a nie jak w przypadku sieci rekurencyjnych każde słowo oddzielnie. Idea ta została zaadaptowana również do innego rodzaju problemów, poprzez zamianę Transformera na Perceiver [80, 80]. Ponieważ celem niniejszej pracy jest ogólna klasyfikacja zachowań postaci ludzkiej, oraz sprawdzenie jak w zależności od ustawień kamery zmienia się rozpoznawalność poszczególnych akcji, do rozpoznawania wybrano tylko podstawowe wersje sieci - jednowymiarowy CNN, oraz wspomniane dwa warianty sieci LSTM. Wprowadzenie wymienionych powyżej modyfikacji pozostaje tematem do rozważenia w przyszłych badaniach.

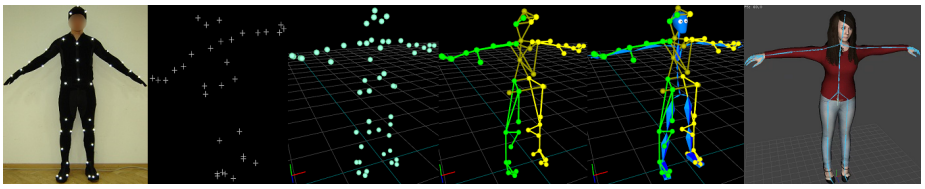
4 Przygotowanie bazy danych

4.1 Motion Capture

Technologia przechwytywania ruchu (ang. Motion Capture, MoCap) [81] dzięki swojej niezwykłej precyzji wykorzystywana jest nie tylko do tworzenia realistycznych animacji komputerowych, ale również znalazła zastosowanie w takich dziedzinach jak medycyna, sport, robotyka, czy biomechanika [82]. Istnieje wiele różnych systemów przechwytywania ruchu, jednakże obecnie za złoty standard uznaje się optyczne systemy MoCap. Systemy te składają się z kilku lub nawet kilkudziesięciu specjalnych kamer emitujących światło podczerwone. Światło to następnie odbija się od specjalnych znaczników znajdujących się na ciele aktora i wraca do kamery, która w ten sposób śledzi ruch znacznika. Rozmieszczenie markerów w dużej mierze zależy od celu nagrania. W przypadku nagrań na potrzeby animacji aktorzy ubierają się w specjalny kombinezon do którego następnie przymocowywane są znaczniki (rys. 7a). Do celów medycznych, gdzie wymagana jest dużo większa precyzja i dokładność, markery są przymocowywane bezpośrednio do skóry pacjenta (rys. 8a). W ten sposób można uniknąć wszelkich drobnych błędów wynikających z ruchu materiału. Następnie niezależnie od specyfiki nagrań, przebieg sesji wygląda podobnie. Kamery przechwytują ruch wszystkich markerów znajdujących się na scenie (rys. 7b i 8b). Jeśli dany znacznik został zarejestrowany przez minimum 2 lub więcej kamer (ich dokładną liczba zależy od łącznej liczby kamer w systemie i jest ustalana przez operatora systemu) zostaje on zrekonstruowany w trójwymiarowej przestrzeni (rys. 7c i 8c). Zrekonstruowane markery tworzą chmurę punktów, którym następnie zostaje nadana nazwa zgodna z wykorzystywanym schematem(rys. 7d i 8d). W przypadku nagrań animacyjnych kolejnym krokiem jest dopasowanie uproszczonego szkieletu i zmapowanie ruchu znaczników na ruch szkieletu (rys. 7e). Następnie ruch ten można wyeksportować do jednego z popularnych formatów, np. fbx i podpiąć do dowolnego modelu 3D (rys. 7f). Nagrania na potrzeby medyczne przetwarzane są w trochę inny sposób - na podstawie zarejestrowanego ruchu, oraz danych zmierzonych przez operatora (np. długości kończyn, waga czy wzrost pacjenta) wyliczane są środki oraz osie poszczególnych stawów (rys. 8e). Następnie na podstawie tych danych wyliczane są parametry kinetyczne (jak kąty w poszczególnych stawach rys.8f) oraz kinematyczne (siła, moc, oraz momenty w poszczególnych stawach). Dodatkowo, systemy stworzone na potrzeby medyczne najczęściej wyposażone są w dodatkowe podsystemy, pozwalające na synchroniczny pomiar dodatkowych parametrów takich jak napięcia mięśniowe, dane z płyt dynamometrycznych, czy referencyjny materiał wideo. Zarówno w przypadku nagrań animacyjnych jak i medycznych, może dojść do sytuacji, w której żadna z kamer nie zarejestrowała znacznika (np. osoba zasłoniła marker ręką). Dlatego też, po etapie dopasowania modelu operator systemu przegląda dane nagranie i uzupełnia wszystkie brakujące trajektorie oraz poprawia ewentualne błędy wynikające ze złego dopasowania.

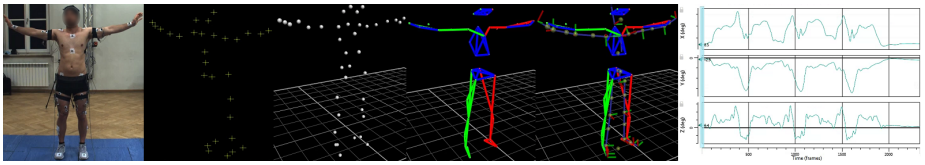
4.2 System akwizycji ruchu

Dane wykorzystane w niniejszej pracy pochodzą z laboratorium Human Motion Laboratory (HML) mieszczącego się w Centrum Badawczo Rozwojowym Polsko Japońskiej Akademii Technik Komputerowych (CBR-PJATK) w Byto-



(**a**) (**b**) (**c**) (**d**) (**e**) (**f**)

Rysunek 7: Etapy akwizycji danych na potrzeby animacji komputerowej: aktor (**a**), widok z kamery (**b**), zrekonstruowana chmura markerów (**c**), dopasowany model (**d**), dopasowany szkielet (**e**), ruch przeniesiony na model 3D (**f**)



(**a**) (**b**) (**c**) (**d**) (**e**) (**f**)

Rysunek 8: Etapy akwizycji danych na potrzeby analizy medycznej: pacjent (**a**), widok z kamery (**b**), chmura markerów (**c**), dopasowany model (**d**), środki oraz osie stawów (**e**), przykładowy wykres dla zgięcia w stawie biodrowym (**f**)

miu. [2]. Laboratorium rozpoczęło działalność w maju 2010 roku. Początkowo wyposażone było w 10 kamer T-40 firmy Vicon, w kolejnych latach było sukcesywnie rozbudowywane. Aktualnie laboratorium wyposażone jest łącznie w 30 kamer firmy Vicon trzech różnych typów: wspomniane wcześniej kamery T-40, Bonita 10 oraz Vantage V5. Podstawowe parametry tych kamer, oraz różnice pomiędzy nimi zostały przedstawione w tabeli 2.

Tabela 2: Podstawowe parametry wykorzystanych kamer

Model kamery	MX-T40	Bonita10	Vantage V5
Rozdzielczość [MP]	4	1	5
Częstotliwość [HZ]	370 @ 4 MP	250 @ 1 MP	420 @ 5 MP
Ogniskowa [mm]	18	4	8.5
Sensor	CMOS	CMOSIS	CMOS
LEDy	180 nm NIR	780 nm NIR	850 nm (IR)
Ilość LEDów	252	68	22
Kąt widzenia (HxV)°	49.15 x 37.14	70,29 x 70,29	63.5 x 55.1
Wymiary [mm],(HxWxD)	207 x 130 x 75	122 x 80 x 79	166.2 x 125 x 134.1
Waga [kg]	1,8	1	1,6

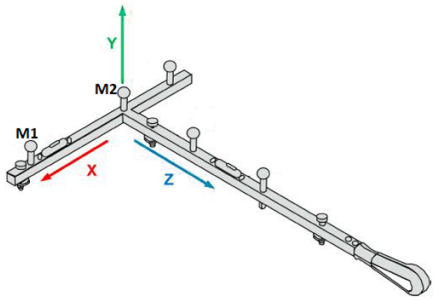
Laboratorium dodatkowo wyposażone jest dwie platformy dynamometryczne firmy Kistler umożliwiające pomiar reakcji podłoża, system do rejestracji

[2] http://bytom.pja.edu.pl/

sEMG firmy Noraxon, oraz 4 kamery wideo Basler Pilot. Dzięki temu w laboratorium możliwa jest rejestracja ruchu zarówno na potrzeby animacyjne jak i medyczne. Nagrania ruchu na potrzeby animacji odbywa się z wykorzystaniem programu Vicon Blade. Podczas tego typu nagrań stosowany jest standardowy schemat składający się z 53 znaczników. W przypadku sesji medycznych wykorzystywane jest drugie z dostępnych oprogramowań - Vicon Nexus. Do nagrań stosuje się standardowy model - Vicon FullBody Plug-In Gait, składający się z 39 znaczników. Model ten dodatkowo wymaga wprowadzenia pomiarów antropometrycznych aktora/pacjenta.

4.2.1 Dokładność danych

Pomimo uznania za złoty standard, optyczne systemy MoCap nie są bezbłędne. Na jakość danych ma wpływ wiele czynników zarówno wewnętrznych jak i zewnętrznych, z których część można modyfikować. Spore znaczenie ma nie tylko ilość i rodzaj wykorzystanych kamer, ale i sposób ich montażu. Kamery zamontowane na statywach mogą wpadać w mikro-wibracje powodowane poruszaniem się na scenie aktorów. W celu ich minimalizacji w laboratorium HML kamery są zamontowane na specjalnym rusztowaniu przymocowanym do ścian. Na jakość danych wpływają też warunki otoczenia takie jak wilgotność czy temperatura. A także ilość oraz rozmiar wykorzystanych znaczników. Jednak największy wpływ ma proces kalibracji kamer MoCap, który polega na odwzorowaniu rzeczywistego położenia kamer na wirtualnej scenie 3D. Proces ten polega na wykonaniu specjalnego nagrania z wykorzystaniem przyrządu kalibracyjnego tzw. T-Frame (rys. 9). Podczas tego nagrania każda z kamer musi zarejestrować określoną liczbę ramek, na których widać wszystkie pięć markerów. Na tej podstawie wyliczane są orientacyjne położenia kamer, oraz błąd re-projekcji - jest to różnica pomiędzy zarejestrowaną pozycją markera a projekcją zrekonstruowanego znacznika na wirtualną kamerę wyrażona w pikselach. Kalibracje systemu uważa się na poprawną, gdy średni błąd reprojekcji wynosi poniżej 0.2 piksela.



Rysunek 9: Przyrząd kalibracyjny T-Frame

Jednakże pomimo zapewnienia stałych warunków otoczenia, oraz perfekcyjnej kalibracji, pozycja zrekonstruowanego nieruchomego znacznika umieszczonego na środku sali zawsze będzie się minimalnie zmieniać. Niemożliwym jest, bowiem uzyskanie błędu re-projekcji na poziomie 0, gdyż każda z kamer obarczona jest dodatkowo wewnętrznym szumem. Szumy występujące w systemie

wykorzystywanym w laboratorium HML zostały szczegółowo zbadane i opisane w artykule [83]. Na potrzeby tych badań wykonano specjalne 9 godzinne nagranie dwóch nieruchomych markerów z przyrządu kalibracyjnego umieszczonego na środku sceny. Podczas nagrania laboratorium było puste - po skalibrowaniu i ustawieniu T-Frame, obsługa opuściła salę a następnie zdalnie uruchomiła i wyłączyła nagranie. Następnie nagranie to było wielokrotnie przetwarzane - markery były rekonstruowane za pomocą różnej liczby kamer. Tak uzyskany zbiór danych był analizowany przy użyciu wariancji Allana [84]. W laboratorium HML występuje przede wszystkim szum biały i losowy, oraz skorelowany prawdopodobnie powiązany z zakłóceniami z zewnątrz.

4.3 Wybór akcji prostych

We wcześniej wspomnianych artykułach, autorzy skupiali się na rozpoznawaniu bardzo dużej liczby akcji. Jednakże to podejście nie ma większego zastosowania w praktyce - w danych pochodzących z systemów monitoringu bardzo rzadko mamy do czynienia z osobami wykonującymi np. pompki, skłony, pajacyki, albo podskoki na jednej nodze. Najczęściej spotykanymi akcjami są chód, stanie w miejscu, obracanie/rozglądanie się, czy schylanie. Dodatkowo z punktu widzenia użyteczności ważnym aspektem jest wykrywanie sytuacji potencjalnie niebezpiecznych jak upadek/leżenie na ziemi, uderzenie albo kopnięcie. Większość z tych akcji została zarejestrowana laboratorium HML w ramach różnych projektów.

4.3.1 Akcja chód

HML posiada bardzo dużą bazę danych różnego rodzaju chodu. Zawiera ona nie tylko chód osób zdrowych, ale również chody pacjentów z różnymi przypadłościami wpływającymi nie tylko na chód, ale i ogólne zdrowie pacjenta, takimi jak:

1. stan po endoprotezoplastyce stawu biodrowego
2. stan po przebytym udarze niedokrwiennym mózgu
3. zwyrodnienie kręgosłupa w odcinku lędźwiowym bądź szyjnym
4. z chorobą Parkinsona, po wszczepieniu neurostymulatora

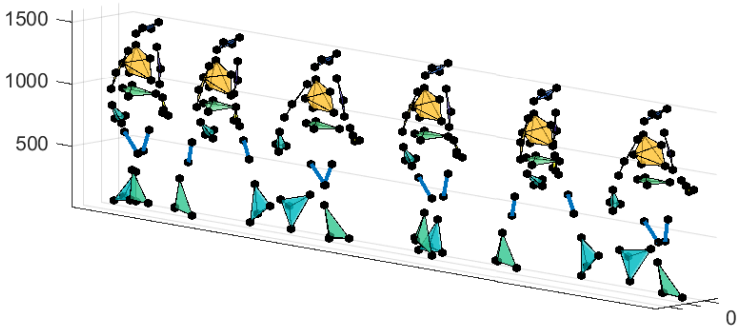
W ramach poszczególnych projektów dodatkowo rejestrowany był też referencyjny chód osób zdrowych. Łącznie zarejestrowano 3347 różnych przejść wykonanych przez 222 osoby. Ponieważ ilość nagrań chodu jest niemal dwukrotnie większa w stosunku do pozostałych akcji, postanowiono podzielić tę grupę na dwie podgrupy - chód osób zdrowych i chód osób chorych. Jednakże podział ten musiał zostać zweryfikowany dla jednej z grup - pacjentów z chorobą Parkinsona. Wynika to ze specyfiki prowadzonych badań [85–88]. Pacjenci w ramach tego projektu wykonywali czterokrotnie kilka ćwiczeń, w tym kilka wariantów chodu. Przy każdym powtórzeniu znajdowali się oni w jednym z poniższych "stanów":

1. bez lekarstw i z wyłączonym neurostymulatorem
2. bez lekarstw i z włączonym neurostymulatorem

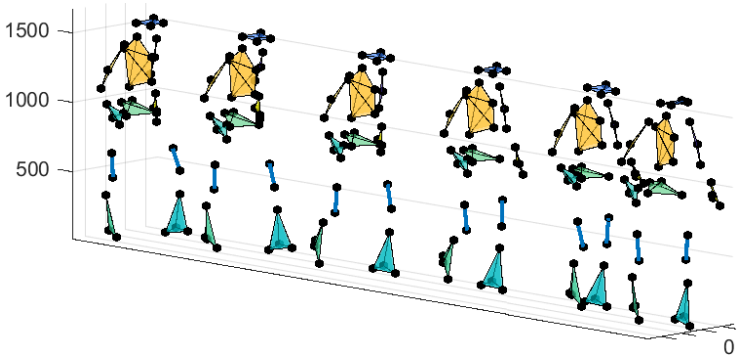
3. z lekarstwami i z wyłączonym neurostymulatorem

4. z lekarstwami i z włączonym neurostymulatorem

Ponieważ zarówno neurostymulator jak i podanie lekarstw znacząco wpływało na zachowanie pacjentów, postanowiono, że nie wszystkie przejścia danego pacjenta znajdą się w grupie - chód osób chorych. O tym czy dane przejście kwalifikowało się do tej grupy decydowała ocena przez lekarza pacjenta w danym stanie. Do oceny lekarz wykorzystywał skalę UPDRS (Ujednolicona Skala Oceny Choroby Parkinsona, ang. Unified. Parkinson's Disease Rating Scale). Otrzymanie wyniku większego niż 1 w kategorii chód, oraz "zamrożenie" chodu klasyfikowało pacjenta do grupy - chód osób chorych. Przykładowy przebieg akcji chód osoby zdrowej oraz chód osoby chorej został przedstawiony kolejno na rysunkach 10 i 11.



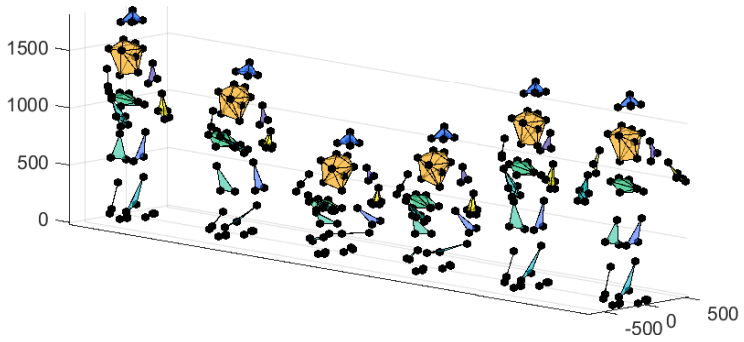
Rysunek 10: Przebieg akcji "Chód osób zdrowych"



Rysunek 11: Przebieg akcji "Chód osób chorych"

4.3.2 Akcja schylanie się

Kolejną licznie reprezentowaną akcją jest - schylanie się/przysiad. Wynika to z faktu, iż na początku każdej sesji pacjent/aktor podlega procesowi kalibracji. Polega on na nagraniu danej osoby podczas wykonywania po kolei ruchu w każdym ze stawów. Następnie na podstawie tego nagrania dopasowywany jest ogólny model oraz tworzony jest unikalny model dla danej osoby z danym ułożeniem markerów. Ponieważ skłon/przysiad jest częścią procedury kalibracyjnej dla danych animacyjnych uzyskano łącznie 1316 nagrań wykonanych przez 307 osób. Przykładowy przebieg tej akcji został przedstawiony na rysunku 12.



Rysunek 12: Przebieg akcji "Schylanie się"

4.3.3 Akcje niebezpieczne

Akcje uznawane za potencjalnie niebezpieczne, czyli uderzenia i kopnięcia zostały zarejestrowane w ramach projektów dotyczących sztuk walk takich jak karate [89] czy taekwondo [90–93]. W przypadku karate zarejestrowano następujące techniki:

- Mae-Geri - podstawowe kopnięcie
- Mawashi-Geri - wysokie kopnięcie
- Ushiro-Mawashi-Geri - wysokie kopnięcie z obrotem
- Gyaku-Zuki - uderzenie

Każda z wymienionych technik została wykonana w jednym z trzech wariantów - uderzenie/kopniak w powietrze (bez przeciwnika), uderzenie/kopniak w tarcze treningową, uderzenie/kopniak w przeciwnika (zarejestrowane wraz z obroną). Nie wszystkie wymienione techniki zostały wykonane przez wszystkich zawodników - mniej zaawansowani mieli problemy z wykonaniem techniki Ushiro-Mawashi-Geri. Szczegółowy opis wraz z ilość wykonanych powtórzeń można znaleźć w artykule [94].

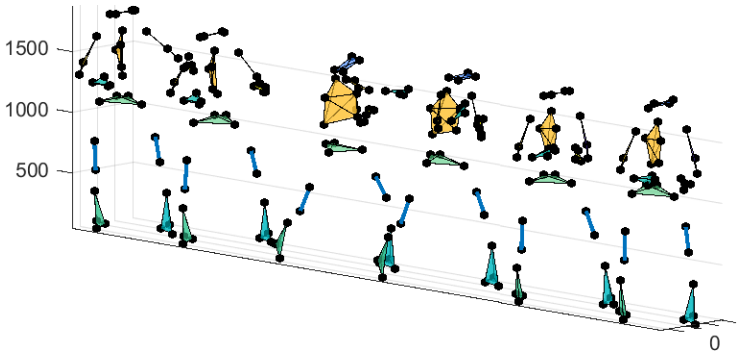
Zawodnicy taekwondo natomiast wykonywali następujące techniki:

- Jirugi - uderzenie proste
- Dollyo - kopnięcie okrężne

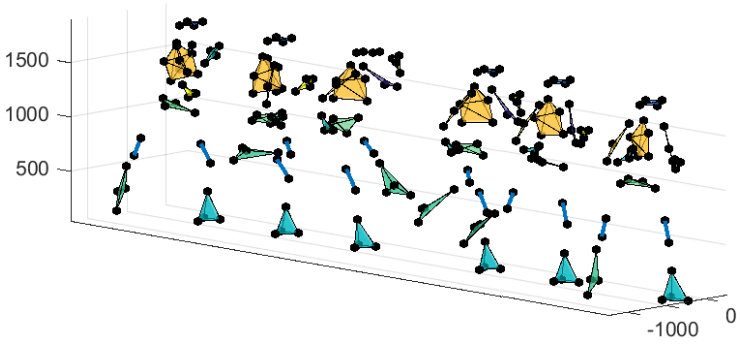
- Yop - kopnięcie boczne
- Ap - kopnięcie frontalne

Podobnie jak w przypadku nagrań z Karate, każda technik została wykonana w kilku wariantach - w powietrze (bez przeciwnika), w piłeczkę tenisową (zawieszoną na lince pod sufitem), w miękką tarczę oraz twardą deskę treningową.

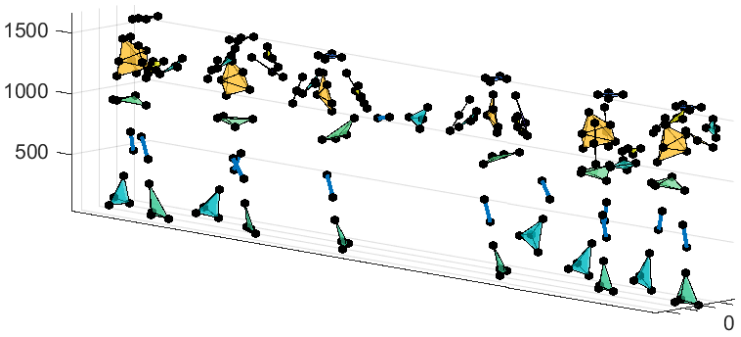
Łącznie w ramach obu projektów nagrano 1305 uderzeń oraz 4444 różnych kopnięć wykonanych przez 62 różne osoby. Ponownie jak w przypadku chodu, ze względu na dysproporcję w ilości nagrań, podjęto decyzję o podziale akcji kopnięcie. Za kryterium podziału wzięto rodzaj kopnięcia - niskie lub wysokie. Dodatkowo kopnięcie wysokie podzielono na dwie podgrupy - proste i boczne. Przebieg utworzonych w ten sposób akcji przedstawiony jest na rysunkach 13, 14, 15 i 16.



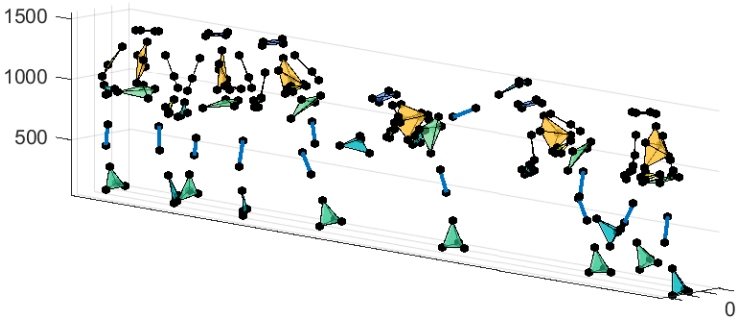
Rysunek 13: Przebieg akcji "Uderzenie"



Rysunek 14: Przebieg akcji "Kopnięcie niskie"



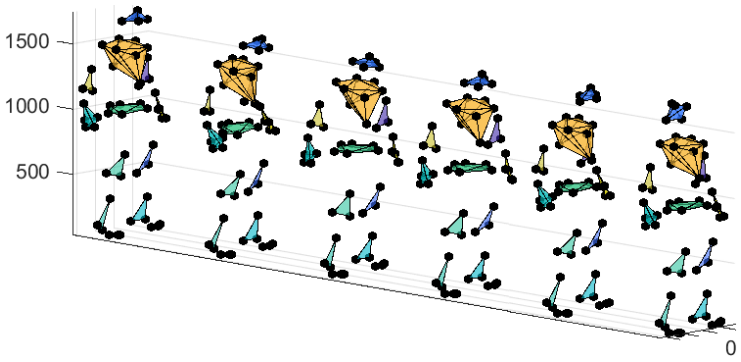
Rysunek 15: Przebieg akcji "Kopnięcie wysokie proste"



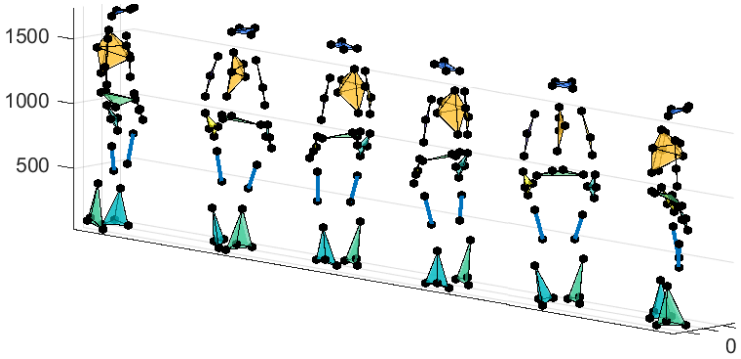
Rysunek 16: Przebieg akcji "Kopnięcie wysokie boczne"

4.3.4 Akcje statyczne

Pozostałe dwie akcje - stanie i obroty, były rejestrowane w ramach różnych mniejszych i większych projektów. Akcja stanie posiada dość silną reprezentację - 1684 nagrań wykonanych przez 81 osób. Niestety akcja obroty/rozglądanie się jest stosunkowo mało liczna. Tylko 625 powtórzeń wykonanych przez 21 osób. Niemniej jest to dość ważna akcja, dlatego postanowiono ją zachować. Przykładowe przebiegi tych akcji przedstawiono kolejno na rysunkach 17 i rys. 18.



Rysunek 17: Przebieg akcji "Stanie"



Rysunek 18: Przebieg akcji "Obroty"

Szczegółowe podsumowanie liczby wszystkich omówionych akcji znajduje się w tabeli 3, wraz ze średnim, minimalnym oraz maksymalnym czasem trwania danej akcji. Natomiast w tabeli 4 znajduje się charakterystyka osób biorących udział w danym nagraniu wraz z podziałem na płeć.

Tabela 3: Wybrane akcje proste

ID	Akcja	Liczba		Czas trwania akcji [s]		
		osób	nagrań	średni	min	maks
Action01	Chód zdrowi	133	1619	4,98	1,47	10,19
Action02	Chód chorzy	89	1728	7,81	2,22	39,22
Action03	Stanie	81	1684	6,21	3,30	11,50
Action04	Obroty	21	625	7,83	3,15	20,70
Action05	Schylanie się	307	1316	1,75	0,48	11,91
Action06	Uderzenie	62	1305	3,68	0,80	12,24
Action07	Kopnięcie niskie	37	1083	3,29	1,55	9,46
Action08	Kopnięcie wysokie proste	62	2006	4,14	2,06	9,13
Action09	Kopnięcie wysoke boczne	62	1355	3,98	1,88	8,73

Tabela 4: Charakterystyka poszczególnych grup

Akcja	Kobiety			Mężczyźni		
	Liczba	Wiek		Liczba	Wiek	
		średni	zakres		średni	zakres
Chód zdrowi	71	28,5	18-60	64	25,5	17-45
Chód chorzy	45	65	36-82	44	60	26-75
Stanie	39	45	20-82	42	45	20-75
Obroty	5	57,5	30-70	16	57	29-73
Schylanie się	73	25,5	23-60	234	25,5	16-49
Uderzenie	23	18	10-31	39	20	10-50
Kopnięcie niskie	13	17	10-31	24	19	10-50
Kopnięcie wysokie proste	23	18	10-31	39	20	10-50
Kopnięcie wysokie boczne	23	18	10-31	39	20	10-50

4.4 Przygotowanie bazy danych 3D

W zależności od potrzeb laboratorium HML dostarcza dane w różnych formatach - szkielet najczęściej w formacie .fbx, pozycje markerów w formacie .csv lub .c3d[3]. Niezależnie od wybranego formatu dane te wymagają dodatkowego przetworzenia. Wynika to ze specyfiki prowadzonych nagrań, wymagań dla poszczególnych projektów, oraz wykorzystania danych pochodzących z dwóch różnych oprogramowań.

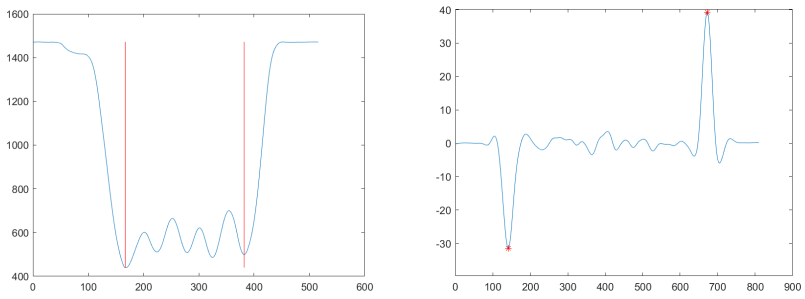
4.4.1 Standaryzacja danych

Przed rozpoczęciem jakiejkolwiek obróbki danych niezbędne było ujednolicenie układów współrzędnych. Domyślnie w programie Blade osie ustawione są następująco: Y - w górę, X - na boki, a Z - wzdłuż sali (układ lewoskrętny). W Nexusie natomiast oś Y i Z są zamienione. Dlatego też, na etapie wczytywania danych układy współrzędnych zostały znormalizowane - przyjęto domyśle ustawienia z Nexusa.

Drugim etapem było usunięcie tzw. T-pozycji, którą pacjent/aktor przyjmuje na początku oraz na końcu nagrania. Ma to na celu ułatwienie późniejszej obróbki danych - przyjęcie tej pozycji sprawia, że w kilku pierwszych i ostatnich ramkach nagrania widoczne są wszystkie markery, co znacząco ułatwia uzupełnienie ewentualnych luk w ich trajektorii. Z punktu widzenia rozpoznawania danej akcji te fragmenty nagrania są zbędne i należy je wyciąć. W tym celu opracowano algorytm analizujący zmianę odległości pomiędzy markerami znajdującymi się na obu dłoniach. Odległość ta jest największa na początku i na końcu gdy aktor przyjmuje T-pozycje rys.19(a). Aby wyciąć T-pozycję na początku nagrania należy znaleźć moment, w którym odległość przestaje maleć. Z kolei na końcu konieczne jest wykrycie momentu, w którym zaczyna rosnąć do maksymalnej wielkości. Algorytm sprowadza się, więc do znalezienia odpowiednio pierwszego lokalnego minimum i ostatniego lokalnego maksimum w pochodnej odległości pomiędzy znacznikami.

Kolejnym etapem przygotowania danych było ujednolicenie ustawienia początkowego pacjentów/aktorów. W większości nagrań osoby podczas wykony-

[3] https://c3d.org/



(a) (b)

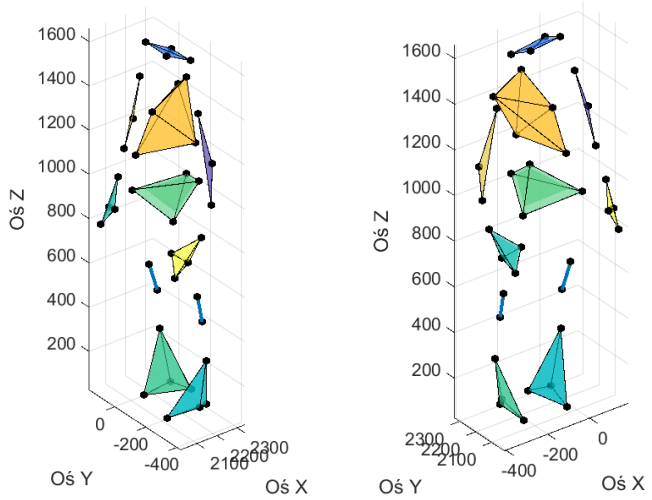
Rysunek 19: Wykresy dla odległości(a) oraz pochodnej odległości(b) dla markerów znajdujących się na dłoniach

wania akcji ustawione były przodem wzdłuż osi Y rys. 20 (b). Jednakże kilka ze wspomnianych wcześniej projektów wymagało, aby pacjent ustawiony był tak, aby patrzył wzdłuż osi X rys. 20 (a). W przypadku tych nagrań, współrzędne markerów musiały zostać prawidłowo obrócone. Wykorzystano do tego proste przekształcenie afiniczne - wszystkie współrzędne *x, y, z* zostały przemnożone przez macierz rotacji wokół osi Z o kąt 90°:

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1' \end{bmatrix} = \begin{bmatrix} cos\alpha & 0 & sin\alpha & 0 \\ 0 & 1 & 0 & 0 \\ -sin\alpha & 0 & cos\alpha & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{17}$$

W przedostatnim etapie przygotowania bazy danych, każdy rodzaj nagrania został poddany innej indywidualnej obróbce. W przypadku danych chodu oraz schylania się wcześniejsze etapy były wystarczające - nie wymagały one dalszego przetwarzania. Nagrania stania oraz obrotów w porównaniu do pozostałych akcji były długie - średnio trwały około minuty, podczas gdy średnia dla pozostałych grup była poniżej 10 sekund. Dlatego też nagrania te zostały podzielone na krótsze fragmenty o losowej długości, tak, aby rozkładem przypominały akcje chód. Najwięcej obróbki wymagały nagrania zawierające kopnięcia i uderzenia. Specyfika projektów, w ramach, których nagrywane były te dane wymagała, aby jedno nagranie zawierało trzy osobne powtórzenia danego ciosu. Dlatego też, każde z nagrań należało podzielić na trzy osobne. Ponieważ już na etapie opracowywania specyfikacji nagrań wiedziano o konieczności ich późniejszego podzielenia, nagrania te zawierały specjalne znaczniki. Operatorzy podczas przetwarzania danych zaznaczali na osi czasu moment rozpoczęcia oraz zakończenia wykonywania poszczególnych technik. W celu podzielenia nagrania na trzy osobne pliki należało więc odczytać te znaczniki z pliku .c3d i na ich podstawie dokonać podziału.

Ostatni etap wymagał ujednolicenia schematu omarkerowania. Jak wspomniano wcześniej, oba oprogramowania korzystają z innych modelów bazujących na różnej liczbie znaczników - Blade na 53 markerach rys.21(a), Nexus na 39 rys.21(c). Jednakże znaczna część znaczników jest dla obu schematów jednakowa - np. markery na poszczególnych stawach, oraz w połowie odległości pomiędzy



Rysunek 20: Ustawienie osoby na scenie przed (a) i po (b) rotacji

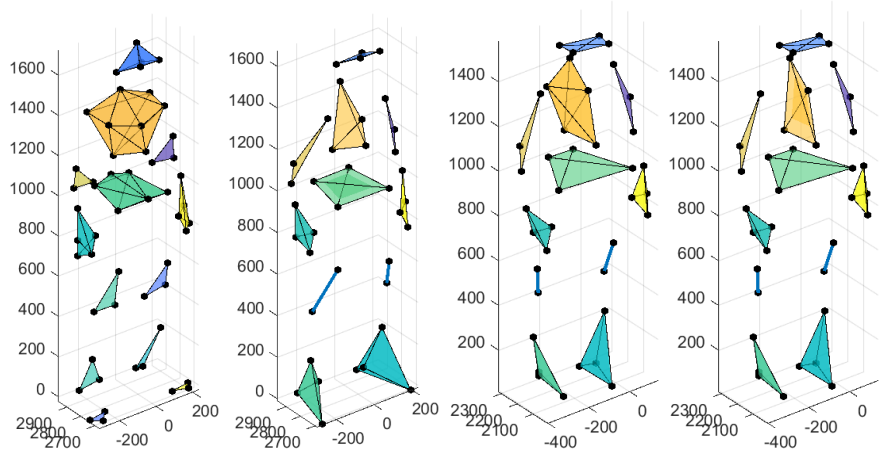
sąsiadującymi stawami. Dlatego też, liczba znaczników została ograniczona do wspólnych 38 markerów rys.21(b,c).

4.4.2 Algorytm podziału na podzbiory

W celu sprawdzenia poprawności modelu, oraz tego czy nie ulega on nadmiernemu dopasowaniu do danych zwykle stosuje się k-krotną walidację krzyżową. Zbiór danych dzieli się na k podzbiorów, z których każdy kolejno służy, jako zbiór testowy, podczas gdy wszystkie pozostałe podzbiory razem stanowią zbiór uczący. W przypadku większości danych podział ten zwykle jest losowy. Jednakże w przypadku omawianej bazy danych zastosowanie całkowicie losowego podziału w znaczącym stopniu zafałszowałoby wyniki. Należy wziąć pod uwagę, że dana osoba powtarzała dany ruch kilkukrotnie, więc istnieje bardzo duże prawdopodobieństwo, iż ruch danej osoby znajdzie się zarówno w zbiorze uczącym jak i testowym. Dlatego też zastosowano specjalny algorytm podziału bazy danych na k podgrup:

1. Utworzenie listy osób wraz z informacją ile razy dana osoba wykonała daną akcję
2. Losowe podzielenie listy osób na k podzbiorów
3. Zsumowanie liczby nagrań w każdym z utworzonych podzbiorów
4. Procentowe porównanie liczby nagrań w każdym podzbiorze
5. W przypadku zbyt dużej dysproporcji podzbiorów powrót do punktu 2

Na rysunku 22 przedstawiono schemat blokowy omawianego algorytmu.



Rysunek 21: Schemat omarkerowania przed i po ujednoliceniu liczby znaczników dla Blade (a,b) i Nexusa (c,d)

Algorytm ten należy powtórzyć osobno dla każdej akcji. Tak uzyskane podzbiory łączymy w k zbiorów. Ponieważ niemożliwym jest uzyskanie identycznej liczby akcji w każdym z podzbiorów, wprowadzono margines błędu. W celu oceny modeli zastosowanych w niniejszej pracy zdecydowano się zastosować 5-krotną walidację krzyżową z 4% marginesem błędu (w każdej z podgrup liczba nagrań nie może być mniejsza niż 18% i nie większa niż 22% wszystkich nagrań). Pozwoliło to na utworzenie pięciu podzbiorów. W tabeli 5 przedstawiono sumaryczne podsumowanie losowo dobranych podzbiorów, które zostały wykorzystane do przeprowadzenia wszystkich zaplanowanych eksperymentów.

Tabela 5: Podsumowanie podzbiorów

Akcja	Liczba nagrań w podzbiorze				
	I	II	III	IV	V
Chód osób zdrowych	325	326	325	328	315
Chód osób chorych	351	346	340	361	330
Stanie	350	330	324	350	330
Obroty	128	126	121	125	125
Schylanie się	261	256	260	273	260
Uderzenie	263	267	255	251	269
Kopnięcie niskie	212	206	217	217	226
Kopnięcie wysokie proste	404	392	403	404	401
Kopnięcie wysokie boczne	274	272	266	259	278

Rysunek 22: Schemat blokowy algorytmu podziału na k-podzbiorów walidacyjnych (treningowych/testowych)

4.5 Przygotowanie danych 2D

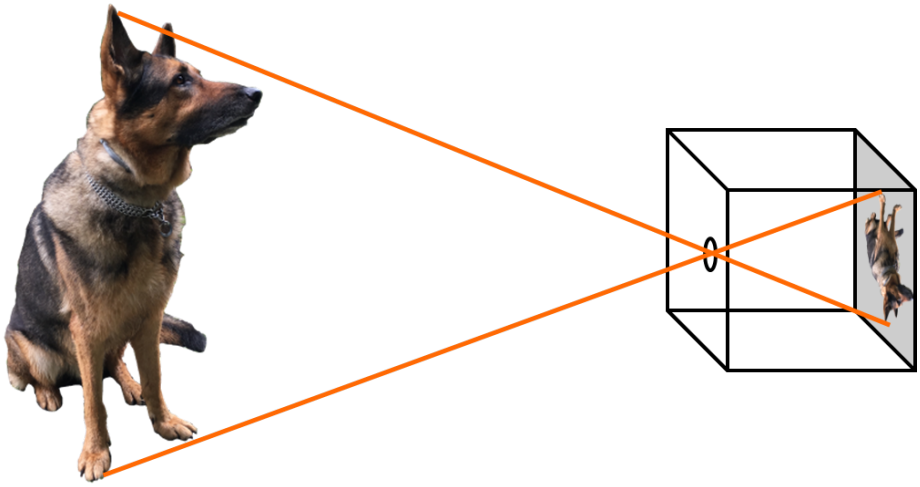
Przygotowanie bazy 2D można podzielić na trzy etapy:

- utworzenie kamery, oraz ustawienie jej parametrów wewnętrznych
- ustawienie pozycji kamery oraz rotacji
- projekcja perspektywiczna

Pierwszy z tych etapów wykonywany jest tylko raz - parametry kamery pozostają stałe niezależnie od jej lokalizacji. Pozostałe dwa powtarzane są kolejno dla wszystkich wybranych położeń wirtualnej kamery.

4.5.1 Wirtualna kamera

Zadaniem wirtualnej kamery jest imitacja rzeczywistej. Jest to zadanie, które głównie polega na symulowaniu sposobu, w jaki światło przemieszcza się w przestrzeni i oddziałuje z obiektami. Istnieje wiele różnych modeli, które w większym lub mniejszym stopniu symulują zachowanie kamer oraz aparatów. W niniejszej pracy zastosowano najprostszy a zarazem najczęściej stosowany model do odwzorowania obiektów z przestrzeni trójwymiarowej na dwuwymiarową - model kamery otworkowej (ang. pinhole camera). Obrazy tworzone za pomocą rzeczywistej kamery otworkowej są do góry nogami. Dzieje się tak dlatego, iż płaszczyzna projekcji znajduje się poza środkiem rzutu (rys. 23). W wirtualnej rzeczywistości, w odróżnieniu do prawdziwego świata, sytuacji tej można bardzo łatwo uniknąć - wystarczy umieścić płaszczyznę rzutowania przed otworem kamery.



Rysunek 23: Schemat kamery otworkowej.

W przypadku wirtualnych kamer konieczne jest też określenie granic widoczności kamery. Granice te wyznacza się za pomocą dwóch płaszczyzn równoległych do płaszczyzny obrazu - bliższej i dalszej. Płaszczyzny te określają które elementy sceny będą odwzorowane na obrazie (rys. 24). Odległość pomiędzy kamerą a płaszczyzną bliższą nazywamy *Znear* a dalszą *Zfar*. W grafice komputerowej bardzo często płaszczyzna bliższa jest tożsama z płaszczyzną obrazu. Niezależnie jednak od umiejscowienia płaszczyzny obrazu musimy wyznaczyć na niej pewien obszar tzw. płótno, na który będzie rzutowany obraz sceny. Współrzędne płótna są niezbędne w celu określenia czy punkt rzutowany na płaszczyznę obrazu będzie widoczny w kamerze czy nie. Współrzędne te są zależne od wielkości płótna, którą można obliczyć za pomocą następującego wzoru:

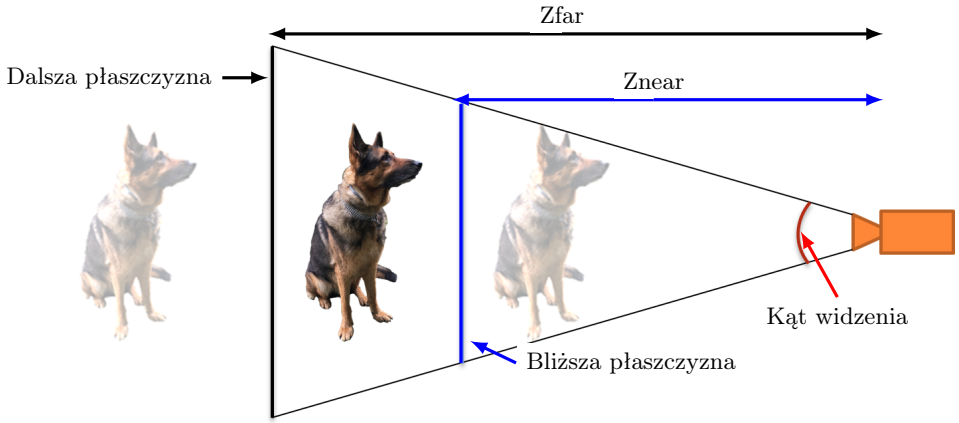
$$Cs = 2 \times tan(\alpha/2) \times Znear \tag{18}$$

Gdzie Cs to rozmiar płótna (z ang. Canvas Size), a α to kąt widzenia kamery. Równanie to jest prawdziwe tylko w przypadku kwadratowego płótna, gdy kąt widzenia pionowy i poziomy są jednakowe. W niniejszej pracy, rozpatrywany jest właśnie taki przypadek. Znając rozmiar płótna, obliczenie jego położenia a w szczególności lewego dolnego i prawego górnego rogu jest proste, biorąc pod uwagę, że jest ono wyśrodkowane na początku układu współrzędnych płaszczyzny obrazu. Należy podzielić rozmiar płótna na 2 i określić znak współrzędnej na podstawie położenia rogu względem układu współrzędnych:

$$TR = \begin{bmatrix} \frac{Cs}{2} \\ \frac{Cs}{2} \end{bmatrix} \tag{19}$$

$$BL = \begin{bmatrix} -\frac{Cs}{2} \\ -\frac{Cs}{2} \end{bmatrix} \tag{20}$$

Gdzie TR to współrzędne prawego górnego wierzchołka (z ang. Top Right), a BL dolnego lewego.



Rysunek 24: Widok boczny parametrów wirtualnej kamery

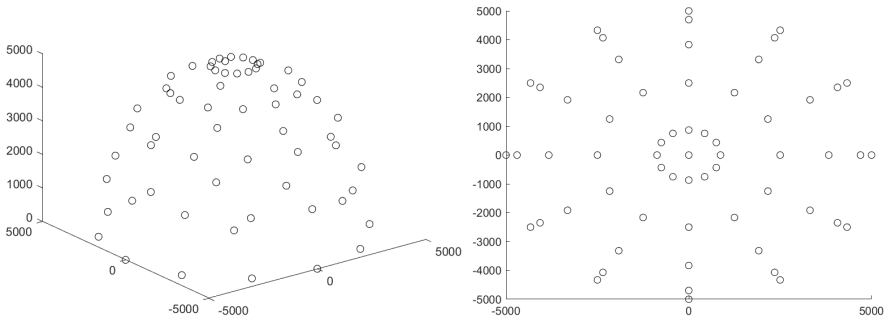
4.5.2 Lokalizacje wirtualnej kamery na scenie

Kolejnym krokiem po utworzeniu kamery i ustawieniu jej parametrów, jest określenie jej położenia na wirtualnej scenie. Położenie kamery określane jest przez zmienne ciągłe które na potrzeby eksperymentów zostały zdyskretyzowane. W tym celu utworzono nad sceną początkową kopułę o średnicy 10 metrów. Wartość została dobrana tak, aby cała osoba była widoczna w kamerze podczas wykonywania danej akcji. Lokalizacje punktów na kopule zostały wyznaczone za pomocą wzorów:

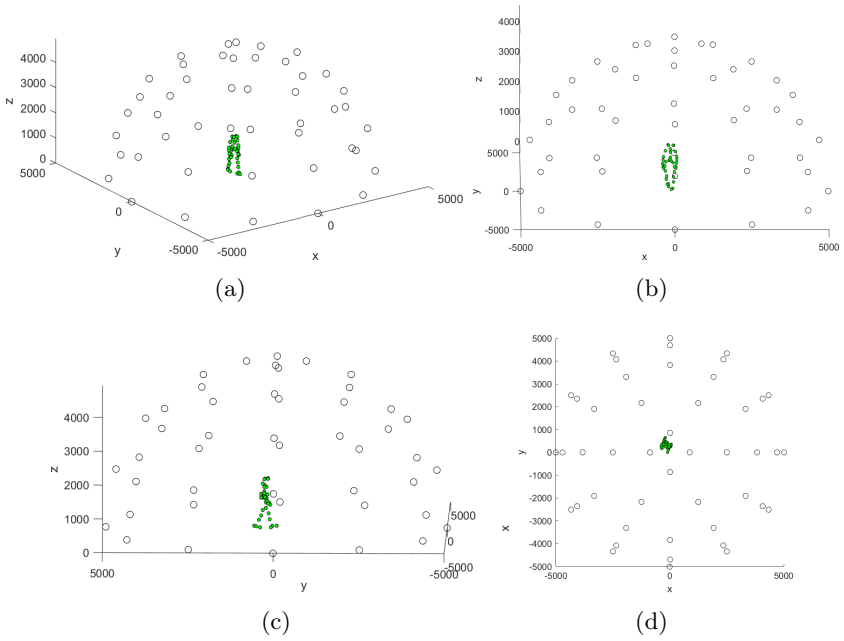
$$x = r \cdot cos(\alpha) \cdot cos(\beta) \tag{21}$$
$$y = r \cdot sin(\alpha) \cdot cos(\beta) \tag{22}$$
$$z = r \cdot sin(\beta) \tag{23}$$

Gdzie r to promień kopuły, α kąt obrotu w płaszczyźnie horyzontalnej, a β kąt obrotu w płaszczyźnie wertykalnej. Wartości obrotów poszczególnych kątów zostały ustalone kolejno na 30°dla kąta α, oraz 20 °dla kąta β. Wizualizacja utworzonych w ten sposób punktów została przedstawiona na rysunku 25.

Ponieważ im bliżej szczytu kopuły tym większe jestb zagęszczenie punktów, ich liczba musiała zostać ograniczona. Ma to na celu utworzenie jak najbardziej różnorodnych baz danych 2D, co w przypadku punktów znajdujących się tak blisko siebie nie było by możliwe. Ostatecznie liczbę lokalizacji wirtualnej kamery ograniczono do 46 różnych pozycji. Ich lokalizacja przedstawiona została na rysunku 26.



Rysunek 25: Wizualizacja wszystkich punktów na kopule, widok z perspektywy (a) oraz z góry (b).



Rysunek 26: Wizualizacja wybranych lokalizacji wirtualnej kamery na przykładzie akcji chód. Widok: perspektywa (a), z przodu (b), z boki (c), z góry (d).

4.5.3 Projekcja perspektywiczna

Ostatnim krokiem w tworzeniu baz danych 2D jest projekcja perspektywiczna, dla każdego położenia wirtualnej kamery. Podczas projekcji wirtualna kamera zawsze skierowana była na środek sceny. Projekcja ta sprowadza się do ciągu przekształceń. W pierwszej kolejności należy znaleźć macierz kamery,

następnie przeliczyć koordynaty wszystkich punktów z układu globalnego na układ kamery, a na końcu dokonać projekcji. Do zdefiniowania macierzy kamery należy określić jej pozycję na scenie, kierunek, w którym patrzy, oraz wektory skierowane w prawo oraz w górę od kamery.

Wyznaczenie pozycji kamery na scenie zostało omówione powyżej w podrozdziale 4.5.2. Wektor kierunkowy kamery wyznaczamy za pomocą następującego wzoru:

$$D = normalize(target - camPos) \tag{24}$$

gdzie *target* to punkt, na który ma patrzeć kamera, a *camPos* to pozycja kamery na scenie. Podczas tworzenia bazy 2D, wirtualna kamera zawsze patrzyła na środek sceny.

Kolejnym wektorem jest wektor skierowany w prawo, który reprezentuje dodatnią oś X przestrzeni kamery. Wyznacza się go za pomocą iloczynu wektorowego wektora kierunkowego, oraz wektora wskazującego w górę w przestrzeni świata (Up):

$$R = normalize(D \times Up) \tag{25}$$

Ponieważ wynikiem iloczynu krzyżowego wektorów, jest wektor prostopadły do obu wektorów, otrzymany w ten sposób wektor wskazuje w kierunku dodatniej osi X.

Ostatni wektor obliczany jest ponownie za pomocą iloczynu wektorowego, tym razem pomiędzy wektorem kierunkowym oraz prawym:

$$U = normalize(R \times D) \tag{26}$$

Wszystkie trzy powyższe wektory wraz z pozycją kamery tworzą macierz przekształcenia współrzędnych:

$$CM = \begin{bmatrix} R_x & R_y & R_z & 0 \\ U_x & U_y & U_z & 0 \\ D_x & D_y & D_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} 1 & 0 & 0 & -P_x \\ 0 & 1 & 0 & -P_y \\ 0 & 0 & 1 & -P_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{27}$$

gdzie R_x, R_y i R_z to składowe *x,y,z* wektora prawego kamery, U_x, U_y i U_z to składowe *x,y,z* wektora pionowego, D_x, D_y i D_z to składowe *x,y,z* wektora kierunkowego, a P_x, P_y i P_z to pozycja *x,y,z* kamery.

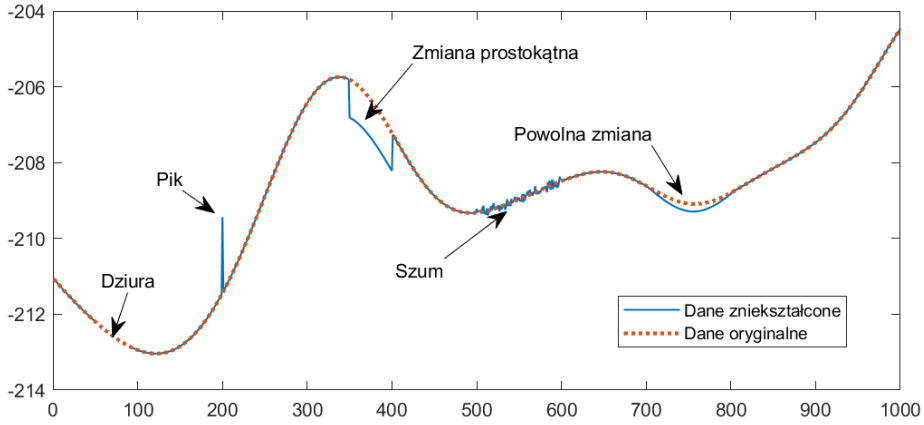
Macierz ta wyliczana jest kolejno dla każdej lokalizacji wirtualnej kamery. Następnie wszystkie współrzędne markerów zostają przekształcone z przestrzeni globalnej na przestrzeń kamery, poprzez przemnożenie ich przez utworzoną macierz. Ostatnim krokiem jest najprostsza projekcja perspektywiczna, przeprowadzana dla każdej pozycji kamery.

Każdy z omówionych wcześniej 5 podzbiorów danych, został poddany rzutowaniu perspektywicznemu kolejno dla każdej z 46 wybranych lokalizacji wirtualnej kamery. Pozwoliło to na utworzenie bazy danych 2D, która została wykorzystana do dalszych testów. Na rysunku 27 przedstawiono przykładowe rezultaty projekcji perspektywicznej, dla trzech losowych pozycji wirtualnej kamery. Pozycje kamery zostały zaznaczone na czerwono. Dodatkowo rzuty te w pełnej rozdzielczości są dostępne w Dodatku A.

4.6 Augmentacja danych

Uczenie maszynowe, a zwłaszcza uczenie głębokie wymaga dużych zbiorów danych. Proces przygotowania - zarejestrowanie, obróbka i etykietowanie, zajmują spore ilości czasu i bywają kosztowne. Dlatego też, w przypadku mniejszych baz danych stosuje się różne metody w celu zwiększenia liczby próbek niewielkim kosztem pracy. W przypadku obrazów najczęściej dodaje się proste transformacje (przesunięcie, skalowanie, obrót, czy odbicie), albo dodanie drobnych szumów, czy zmiany w nasyceniu koloru lub jasności. W przypadku sygnałów można zastosować podobne metody, wymagają one jednak dostosowania do charakterystyki danych. Zniekształcenia te można podzielić na dwie grupy lokalne, oraz globalne.

Przekształcenia na poziomie lokalnym obejmują pojedyncze zmiany w trajektorii markera. Dodatkowo mogą, lecz nie muszą być skorelowane z innymi markerami. Wśród tego rodzaju zniekształceń można wyróżnić: luki/dziury w trajektorii, pojedyncze piki, szumy oraz zmiany skokowe i powolne (rys. 28). Wszystkie te zniekształcenia występują naturalnie w danych MoCap ([95]) i są usuwane przez operatora systemu na etapie oczyszczania danych. Ponieważ dane wykorzystane w niniejszej pracy zostały oczyszczone, zniekształcenia te należałoby dodać sztucznie. Wymagałoby to określenia parametrów dla każdego zakłócenia. Dla pojedynczych pików należy określić ich maksymalną wysokość, oraz częstotliwość występowania. W przypadku szumu poza długością jego występowania, należy dodatkowo określić jego rodzaj np. Gaussowski. W przypadku luk trzeba też rozważyć różne metody ich uzupełniania, od prostej interpolacji do bardziej skomplikowanych metod bazujących na sieciach neuronowych czy drzewach decyzyjnych ([96]). W zależności od zastosowanej metody może to prowadzić do powstawania zmian - skokowych, oraz powolnych.



Rysunek 28: Najczęściej spotykane błędy w danych MoCap

Ponieważ zmiany te dotyczą pojedynczych trajektorii markera, lub nawet jednej z jego składowych, konieczne jest również określenie zakresu tych zmian. Ilu markerów w danym nagraniu mają dotyczyć? Czy jednocześnie ma występować jedno, dwa, czy wszystkie zakłócenia? Ile procent nagrania ma być zniekształcone? Zagadnienie to stanowi osobny obszar badawczy, który może zostać

rozważony w przyszłości. Dlatego też postanowniono nie wprowadzać tego typu zmian.

Przekształcenia globalne obejmują wszystkie znaczniki w danym nagraniu. Do tego rodzaju przekształceń zaliczamy m.in obrócenie czy przesunięcie aktora na scenie, odbicie lustrzane, rozmycie danych poprzez zastosowanie filtru dolnoprzepustowego, czy przycięcie nagrania. Część z tych przekształceń naturalnie występuje już w danych. Aktorzy podczas wykonywania akcji stali w różnych częściach sceny, akcja typu chód zaczynała i kończyła się w innym miejscu dając tym samym efekt przycięcia. Natomiast wprowadzenie rotacji aktora z punktu widzenia omawianych eksperymentów jest niepotrzebne - widok danej akcji pod innym kątem umożliwia rotacja kamery.

Augmentację stosuje się przede wszystkim, gdy zbiór danych jest zbyt mały. Baza danych wykorzystywana w niniejszej pracy jest dość obszerna, a przeprowadzone testy nie wykazały, aby którakolwiek z metod prowadziła do zbytniego dopasowania się do danych. Dlatego też zrezygnowano z tego etapu przygotowywania danych. Warto natomiast rozważyć analizę wpływu poszczególnych zakłóceń, na jakość rozpoznawania poszczególnych akcji, jednakże nie jest to celem niniejszej pracy.

5 Klasyfikacja zachowań postaci ludzkiej w przestrzeni trójwymiarowej

5.1 Sformułowanie zadania

Eksperymenty zaproponowane w niniejszym rozdziale miały na celu dwie rzeczy. Pierwszą z nich była analiza porównawcza trzech różnych typów głębokich sieci neuronowych - jednowymiarowej sieci CNN, oraz dwóch sieci LSTM (jedno i dwukierunkowej).

Drugim celem było określenie optymalnego rozmiaru wektora wejściowego. W przypadku danych pochodzących z systemu Motion Capture mamy do czynienia z nadreprezentacja danych. Podczas nagrań, pacjent ma na sobie 39 lub 53 znaczniki, w zależności od wybranego schematu. Po ujednoliceniu schematów nadal pozostaje 38 markerów. Znaczna ich część znajduje się na tych samych stawach lub segmentach ciała. W przypadku danych z rzeczywistych kamer bardzo trudno jest wyodrębnić i śledzić taką liczbę punktów charakterystycznych dla każdej osoby. Dlatego też liczba danych wejściowych musiała ulec redukcji. Sposób ich eliminacji został opisany w podrozdziale 5.1.1.

Do realizacji tych celów utworzono sieci każdego typu a następnie testowano je za pomocą 5-krotnej walidacji krzyżowej, powtórzonej 5 razy dla każdego z 7 rozmiarów wektora wejściowego. W celu oceny jakości klasyfikacji wykorzystano takie miary jak dokładność (ang. Accuracy, ACC), czułość (ang. True positive rate TPR), częstotliwość błędów (ang. False negative rate FNR), precyzja (ang. Positive predictive value, PPV), czy oczekiwana proporcja błędów (ang. false discovery rate, FDR). Wymienione miary można wyrazić za pomocą następujących wzorów:

$$ACC = \frac{TP + TN}{P} \tag{28}$$

$$TRP = \frac{TP}{TP + FN} \cdot 100\% \tag{29}$$

$$FNR = \frac{FN}{TP + FN} \cdot 100\% \tag{30}$$

$$PPV = \frac{TP}{TP + FP} \cdot 100\% \tag{31}$$

$$FDR = \frac{FP}{TP + FP} \cdot 100\% \tag{32}$$

gdzie TP to liczba prawdziwie poprawnych klasyfikacji (ang. true positive), TN to liczba prawdziwie negatywnych klasyfikacji (ang. true negative, TN), FP to liczba fałszywie pozytywnych klasyfikacji (ang. false positive, FP), FN to liczba fałszywie negatywnych klasyfikacji (ang. false negative, FN), a P to wielkość populacji.

Ze względu na charakter danych, podzbiory wykorzystywane w omawianych eksperymentach nie są zbalansowane - nie możliwym jest uzyskanie takiego samego procentowego podziału w przypadku każdego z k-zbiorów walidacyjnych. W literaturze można spotkać inne miary, takie jak współczynnik korelacji Matthewsa (ang. Matthews Correlation Coefficient MCC) [97], który jest miarą jakości efektywności klasyfikacji dla niezrównoważonych populacji klas. Jego

wartości skalują się od -1 w przypadku braku klasyfikacji do 1 w przypadku dokładnej klasyfikacji. Współczynnik ten oblicza się za pomocą następującego wzoru:

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(FP + FN)(TN + FP)(TN + FN)}} \tag{33}$$

Jednakże nawet takie miary mają mała wartość poznawczą przy mocno niezbalansowanych zbiorach [95]. W związku z czym wprowadzono dwie dodatkowe miary, oraz przeprowadzono pogłębioną analizę dla każdego z zestawów danych. Miary te to kolejno *Poziom Błędu Osoby (PBO)* i *Poziom Błędu Akcji (PBA)*. Ponieważ każda akcja była wykonywana kilka razy przez daną osobę, miara ta ma określać, jaki procent akcji wykonanych przez jedną osobę został błędnie sklasyfikowany, oraz średnią wartość PBO w ramach danej akcji:

$$PBO = \frac{Liczba\ blednych\ klasyfikacji\ osoby}{Wszystkie\ akcje\ osoby} \tag{34}$$

$$PBA = \frac{1}{N} \sum_{i=1}^{N} POB_i \tag{35}$$

Sieci neuronowe, a w szczególności głębokie sieci neuronowe mają tendencje do przeuczania się czy wręcz zapamiętywania konkretnych wzorców. Dlatego też koniecznym jest sprawdzenie, czy w omawianym przypadku sieci są w stanie odpowiednio uogólnić dane i rozdzielić konkretny, charakterystyczny układ śledzonych markerów, od charakterystyki danego ruchu. Jest to główne zadanie wprowadzonych miar, które pozwalają na potencjalne zidentyfikowanie czy sumaryczne błędy sieci nie są zdominowane przez jedną osobę lub jedną akcję.

Wspomniana pogłębiona analiza obejmuje między innymi próby znalezienia przyczyny tych pomyłek w ramach poszczególnych klas poprzez dokładną analizę danych aktorów obejmującą nie tylko analizę sposobu wykonania danej akcji, ale również poprzez porównanie pozostałych danych pacjentów takich jak ich wiek, płeć, czy w przypadku osób chorych nasilenie objawów chorobowych. Sprawdzono również wpływ wykorzystanego oprogramowania, na jakość klasyfikacji. Ostatnim z parametrów branym pod uwagę w dalszej analizie była też data nagrania. Dane nagrywane były na przestrzeni kilku lat, podczas których zmieniało się wyposażenie laboratorium (przybywało kamer, czy zmieniały się wersje oprogramowania), a także doświadczenie pracowników. Wyniki tych analiz znajdują się w rozdziałach 5.3 dla danych trójwymiarowych oraz 6.5 dla danych dwuwymiarowych.

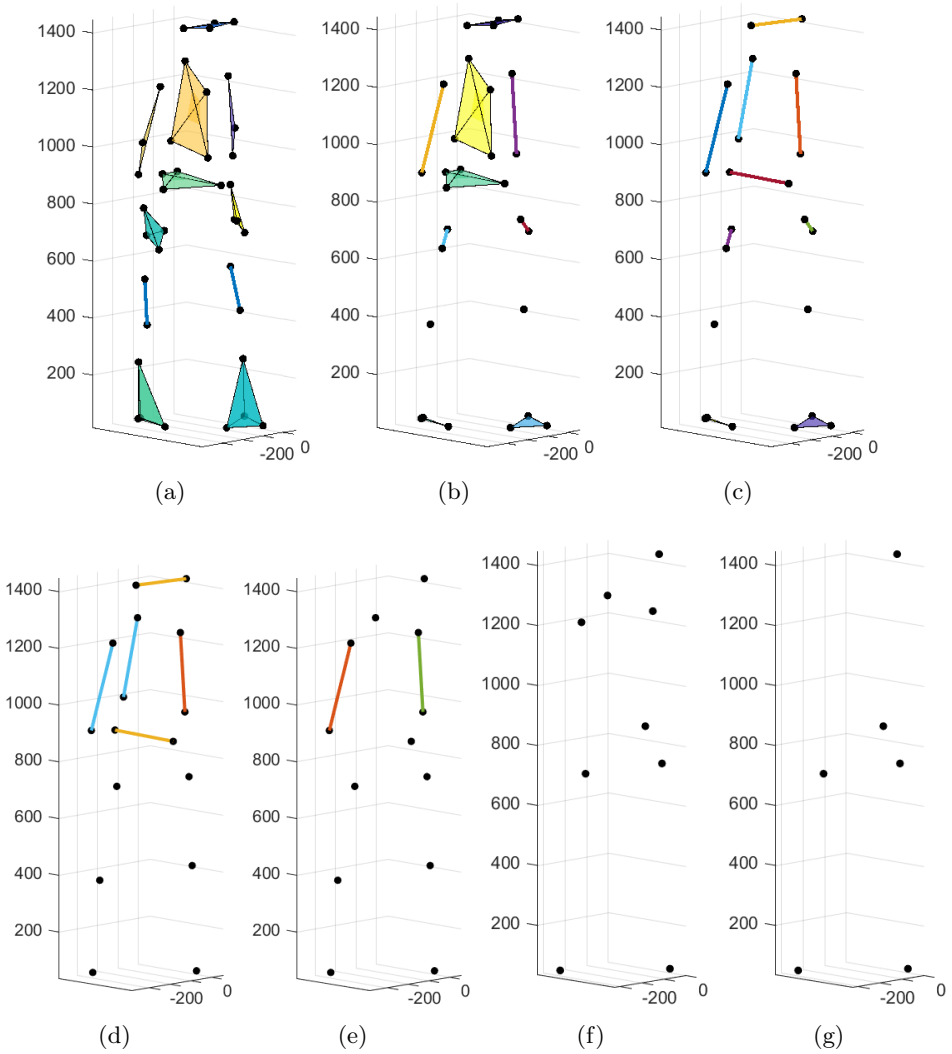
5.1.1 Kryteria redukcji wektora wejściowego

Jak już wcześniej wspomniano, w fizycznych systemach optycznych niemożliwym jest śledzenie tak dużej liczby punktów dla każdej zarejestrowanej osoby. Dlatego też konieczne jest wyznaczenie minimalnej liczby znaczników - takiej, przy której dokładność sieci oraz czułość klasyfikacji pozostają wysokie.

Początkowo jako wektor wejściowy wykorzystywane były trajektorie wszystkich 38 znaczników (29a), a następnie ich ilość ulegała systematycznej redukcji.

Uzyskano w ten sposób siedem różnych wektorów wejściowych, z których każdy zawierał coraz mniejszą liczbę markerów. Eliminacja odbywała się zgodnie z poniższymi kryteriami.

W pierwszym kroku redukcji postanowiono usunąć znaczniki znajdujące się pomiędzy dwoma sąsiadującymi stawami: znaczniki z ramion, przedramion, ud i łydek. Markery te mają ułatwić systemowi rozróżnienie prawej od lewej strony ciała, dlatego ich lokalizacja jest za każdym razem trochę inna. Ich usunięcie ograniczyło wektor wejściowy do 28 markerów (rys. 29 b).



Rysunek 29: Wizualizacja wybranych markerów: (a) 38 znaczników, (b) 28 znaczników, (c) 22 znaczniki, (d) 16 znaczników, (e) 13 znaczników, (f) 9 znaczników, (g) 7 znaczników.

Drugim krokiem była redukcja o połowę markerów znajdujących się na tych samych segmentach ciała: głowie, tułowiu, oraz miednicy. Związane to było głównie z ograniczeniami, jakie wynikają w przypadku wykorzystywania danych z kamer 2D. Ograniczyło to liczbę znaczników do 22 (rys. 29c).

W następnym kroku ograniczono markery znajdujące się bardzo blisko siebie: na stopach oraz dłoniach. Usunięto markery pod paliczkami, na piętach oraz pod palcem środkowym. Pozostawiając tym samym 16 znaczników (rys. 29d).

Dalsza redukcja ponownie obejmowała segmenty - każdy z nich był już tylko reprezentowany przez jeden marker. Z taką sytuacją najczęściej ma miejsce przy danych z kamer wizyjnych. Liczba znaczników została ograniczona do 13 (rys. 29e).

Przedostatnim krokiem była rezygnacja z markerów znajdujących się na kolanach i łokciach. Redukcja ta sprawiła, że każda z kończyn reprezentowana była przez dwa markery odpowiadające początkowi i końcowi kończyny - ramiona-dłonie, miednica-kostki (rys. 29f).

W ostatnim kroku ograniczono liczbę markerów do 7 zostawiając tylko znaczniki na głowie, miednicy oraz końcach kończyn - nadgarstki i kostki. Jest to minimalna liczba markerów jaka może zostać wykorzysta do rozpoznawania akcji. Eliminacja choćby jednego markera więcej sprawi, że dana kończyna lub segment nie będą już w żaden sposób reprezentowane (29g).

5.2 Struktury wybranych sieci razem z opisem hiperparametrów

Po wyborze typu sieci neuronowej, konieczne jest uszczegółowienie jej architektury, oraz ustalenie wszystkich hiperparametrów algorytmu uczącego. Od odpowiedniego doboru tych hiperparametrów zależy, jakość wybranego modelu, oraz to, w jakim stopniu nauczy się on uogólniać dane. Problem ten jest trudny, złożony i bardzo czasochłonny. Na przestrzeni lat różni badacze proponowali rozmaite rozwiązania tego problemu. Jednym z popularniejszych rozwiązań jest zastosowanie optymalizacji Bayesa.

Optymalizacja bayesowska [98] opiera się na algorytmach prawdopodobieństwa warunkowego. Jej zadaniem jest zbadanie przestrzeni wejściowej funkcji celu i znalezienie najbardziej optymalnych parametrów modelu na ograniczonym podzbiorze przestrzeni wejściowej. W odróżnieniu od innych podejść nie polega ona na lokalnym gradiencie i przybliżeniu hesjanu, a wykorzystuje informacje ze wszystkich poprzednich ocen. Dzięki czemu możliwe jest znalezienie lokalnego minimum trudnych niewypukłych funkcji przy stosunkowo niewielkiej liczbie ocen.

Wszystkie eksperymenty omawiane w niniejszej pracy były wykonywane w Matlabie, w którym jest wbudowany optymalizator bayesowski. Do optymalizacji zostały wybrane następujące hiperparametry, wraz z zakresami:

- głębokość sieci (liczba ukrytych warstw od 1 do 5)
- prędkość uczenia się (0.0001 - 0.1)
- przycięcie gradientu (1-inf)

Dodatkowo w przypadku sieci LSTM liczba komórek na poszczególnych warstwach (od 10 do 200). W przypadku sieci CNN liczba (10-100) i rozmiar filtrów

(2-20). Dodatkowo dla sieci CNN przyjęto założenie, że z każdą wartwą liczba filtrów ulega podwojeniu. Optymalizacja została przeprowadzona kilkukrotnie, dla każdego rozmiaru wektora wejściowego, na wstępnym podzbiorze danych.

5.2.1 Sieci LSTM

Dla obu typów sieci LSTM, największy wpływ na ogólną jakość klasyfikacji miał rozmiar wektora wejściowego. Dokładniejsza analiza wpływu wartości hiperparametrów na uzyskane rezultaty wykazała kilka zależności.

Wartości hiperparametrów uzyskanych dla każdego rozmiaru wektora wejściowego oscylowały wokół podobnych wartości. Przykładowo wartość prędkości uczenia się oscylowała od 0.00085 do 0.00110. Co więcej w każdym przypadku optymalna okazywała się sieć z 3 warstwami ukrytymi. Zastosowanie mniejszej liczby warstw powodowało problemy z uogólnieniem danych, z kolei przy pięciu warstwach dochodziło do przeuczenia sieci. W przypadku tego typu sieci dochodziło do błędów w momencie gdy wartość gradientu była nieskończona.

Ponieważ, wpływ rozmiaru wektora wejściowego miał największy wpływ, który został bardziej szczegółowo opisany w rozdziale poświęconym analizie wyników, zdecydowano się uśrednić wartości wszystkich hiperparametrów. Dla jednokierunkowej sieci LSTM wybrano następujące parametry:

- głębokość sieci – 3 warstwy ukryte
- liczba ukrytych jednostek na warstwie – 100/50/50
- prędkość uczenia się – 0.001
- przycięcie gradientu – 10

Dla dwukierunkowej sieci LSTM wybrano następujące parametry:

- głębokość sieci – 3 warstwy ukryte
- liczba ukrytych jednostek na warstwie – 100/100/50
- prędkość uczenia się – 0.001
- przycięcie gradientu – 20

Architektura obu sieci LSTM jest podobna, różną się tylko parametrami uczenia. Schemat obu sieci został przedtsaiony na rysunku 30.

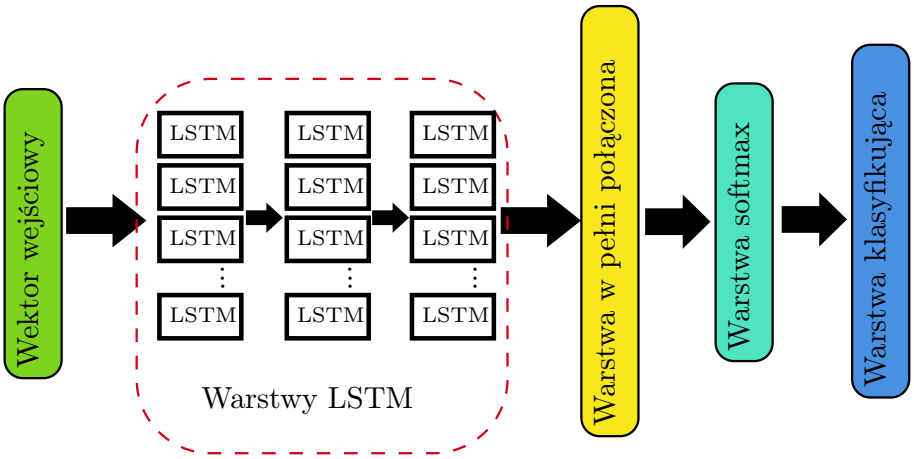
5.2.2 Sieć CNN

W przypadku sieci CNN ponownie rozmiar wektora wejsciowego miał nieco mniejszy wpływ na jakość wyników. Jednakże, podobnie jak w przypadku sieci LSTM do dalszych, bardziej szczegółowych testów utworzono jedną sieć, będącymi uśrednieniem najlepszych wyników:

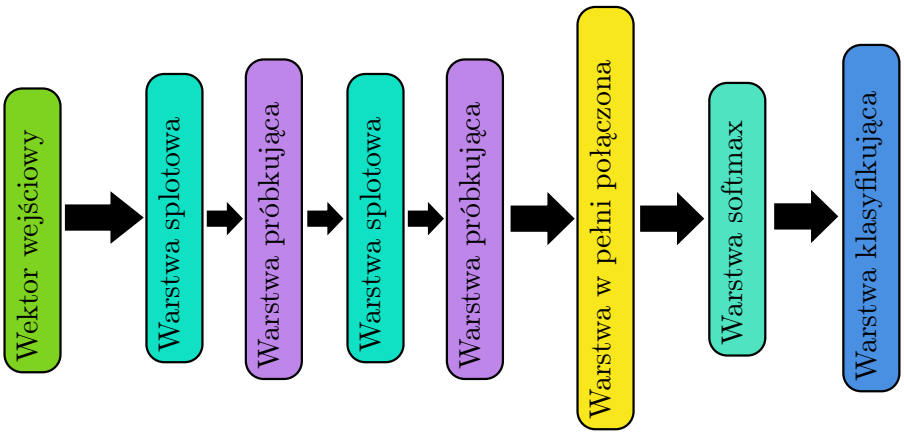
- głębokość sieci – 2 warstwy ukryte
- liczba filtrów na warstwie – 32/64
- rozmiar filtra - 5
- prędkość uczenia się – 0.001

- przycięcie gradientu – inf

Zaprojektowa sieć jest dość generyczna. Jej architektura została przedstawiona na rysunku 31.



Rysunek 30: Schemat zaprojektowanej sieci LSTM

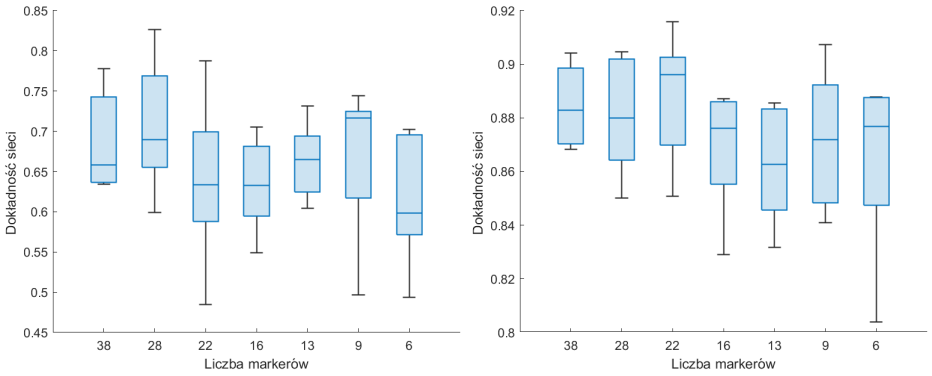


Rysunek 31: Schemat zaprojektowanej sieci CNN

5.3 Klasyfikacja z wykorzystaniem sieci LSTM

5.3.1 Ogólna jakość klasyfikacji

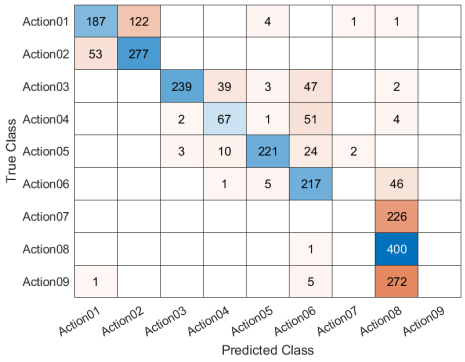
Jak wspomniano wcześniej największy wpływ, na jakość klasyfikacji miała liczba markerów użyta do stworzenia wektora wejściowego. Na rysunku 32 przedstawiono wykresy pudełkowe dla dokładności sieci jedno i dwukierunkowej, dla różnych rozmiarów wektora wejściowego. W przypadku obu sieci możemy zauważyć podobną zależność pomiędzy liczbą markerów na wejściu a jakością sieci. Początkowo redukcja markerów, dzięki eliminacji bardzo podobnych do siebie trajektorii, poprawiała wyniki sieci. Jednakże przekroczenie pewnego progu sprawiło, że nie tylko średnia dokładność sieci zaczęła spadać, ale również jej spójność - zwiększał się zakres wartości dokładności sieci utworzonych dla tych samych danych wejściowych. W przypadku sieci jednokierunkowej wahania wynosiły od 12% do 24%, w przypadku dwukierunkowej od 3,5% do 8,5%. W obu przypadkach spadek liczby markerów powodował zwiększenie zakresu dokładności.



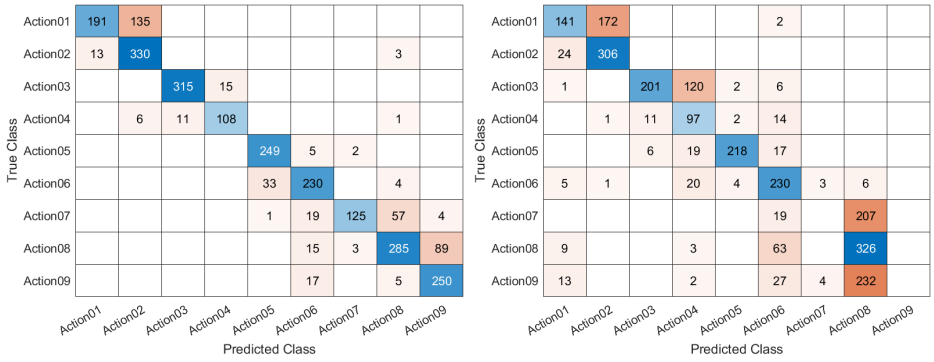
(a) (b)

Rysunek 32: Wykres pudełkowy dla dokładności sieci LSTM (a) jednokierunkowej, (b) dwukierunkowej.

Jednakże sieci jednokierunkowe, niezależnie od rozmiaru wektora wejściowego, uzyskiwały znacząco gorsze rezultaty w porównaniu do sieci dwukierunkowych. Bardzo często wszystkie rodzaje kopnięć klasyfikowały, jako jedno, co w znaczący sposób obniżało dokładność sieci. Dodatkowo dochodziło do licznych pomyłek pomiędzy praktycznie wszystkimi akcjami, co widoczne jest na przykładowych macierzach pomyłek na rysunkach 33 i 34. Dodatkowo na rysunku 34 przedstawiono macierze pomyłek dla tego samego rozmiaru wektora wejściowego oraz danych uczących i testowych. Jakość klasyfikacji pomiędzy tymi sieciami jest skrajnie różna. Dlatego też, sieci jednokierunkowe zostały wyeliminowane z dalszej analizy i nie zostały wzięte pod uwagę w przypadku danych dwuwymiarowych.



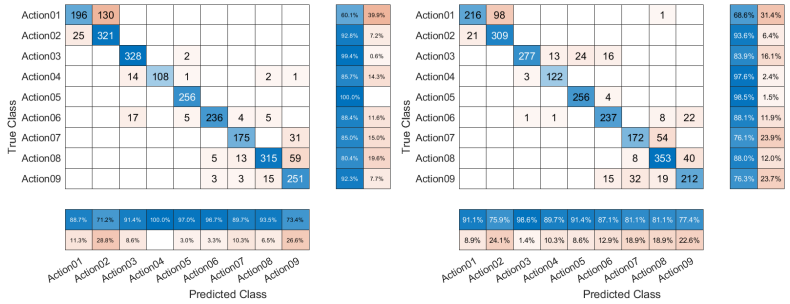
Rysunek 33: Przykładowa macierz pomyłek jednokierunkowej sieci LSTM dla 28 znaczników w wektorze wejściowym



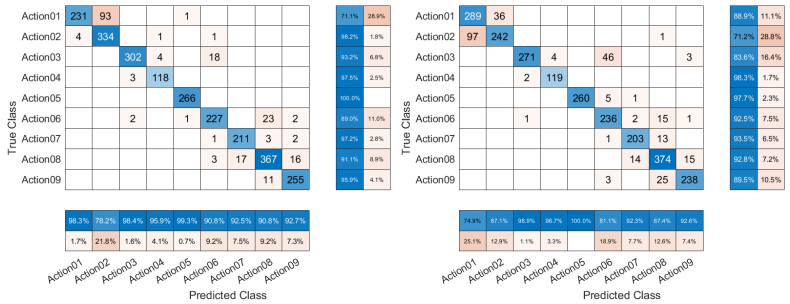
(a) (b)

Rysunek 34: Przykładowe macierze pomyłek jednokierunkowej sieci LSTM dla 22 znaczników w wektorze wejściowym

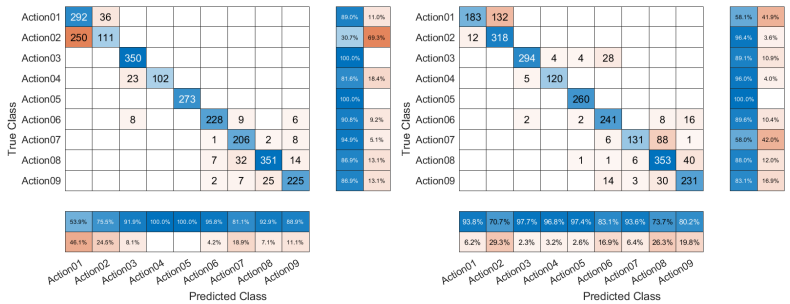
W przypadku dwukierunkowej sieci LSTM można zauważyć grupowanie się pomyłek. Mylone są akcje o podobnej charakterystyce - chody osób zdrowych z chorymi i odwrotnie, czy różne rodzaje kopnięć pomiędzy sobą. Widoczne jest to na rysunku 35, gdzie przedstawiono macierze pomyłek dla każdego rozmiaru wektora wejściowego. Co więcej wraz ze zmniejszaniem się liczby markerów w wektorze wejściowym zwiększała się liczba pomyłek nie tylko pomiędzy podobnymi akcjami, ale również pozostałymi.



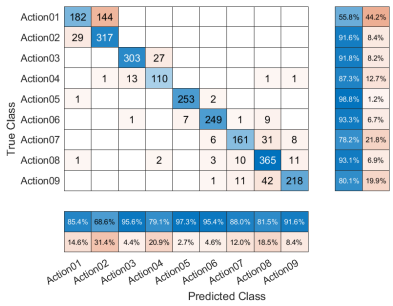
(a) (b)



(c) (d)



(e) (f)

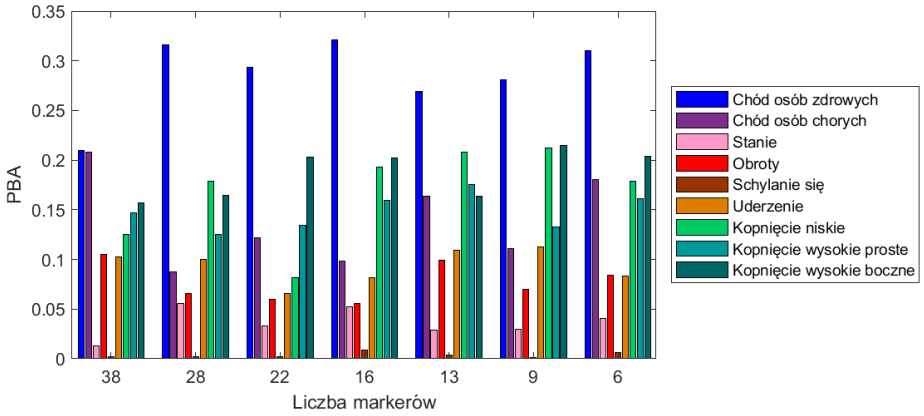


(g)

Rysunek 35: Przykładowe macierze pomyłek dwukierunkowej sieci LSTM dla różnej liczby znaczników w wektorze wejściowym: (a) 38, (b) 28, (c) 22, (d) 16, (e) 13, (f) 9, (g) 6

Na rysunku 36 przedstawiono wykres słupkowy dla wartości PBA dla poszczególnych akcji z uwzględnieniem zmian w liczbie znaczników w wektorze wejściowym. Podobnie jak w przypadku ogólnej, jakości sieci, średnia liczba pomyłek danej osoby w ramach danej akcji rośnie. Mając na uwadze zarówno średnią, jakość sieci, oraz sumaryczną liczbę pomyłek na osobę, najlepsze rezultaty uzyskano dla sieci, która na wejściu miała 22 znaczniki.

Jednakże, należy mieć na uwadze specyfikę wykorzystanych danych - może się okazać, że w ramach jednej akcji, wszystkie powtórzenia tylko 2/3 osób są błędnie klasyfikowane. Dlatego konieczna jest dokładniejsza analiza każdej akcji w połączeniu z danymi aktorów takimi jak wiek, płeć, wzrost, waga czy w przypadku osób chorych nasilenie objawów chorobowych.

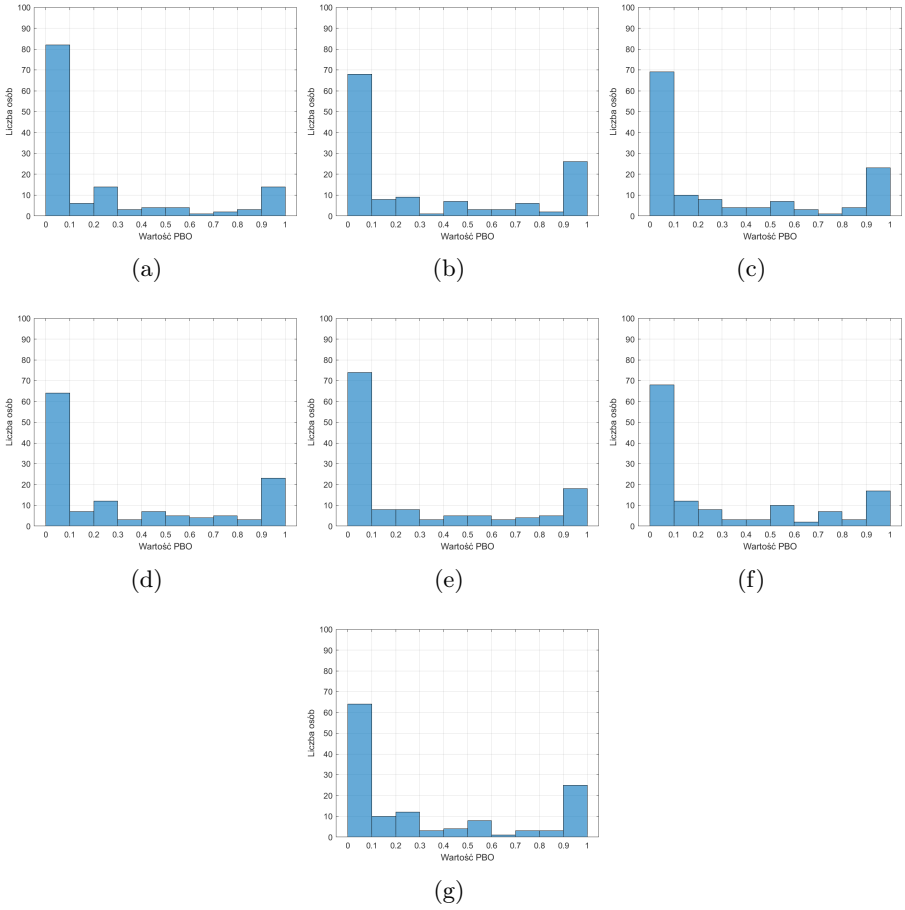


Rysunek 36: Wykresy słupkowe wartości PBA dla różnej liczby markerów w wektorze wejściowym

5.3.2 Akcja chód osób zdrowych

Chód osób zdrowych, poza jednostkowymi przypadkami mylony był z chodem osób chorych. Na rysunku 37 przedstawiono rozkład miary PBO dla poszczególnych rozmiarów wektora wejściowego. Można zauważyć, że wraz ze spadkiem liczby markerów na wejściu liczba osób, których przejścia są zawsze lub prawie zawsze poprawnie klasyfikowane spada. Utrzymuje się ona jednak zawsze, powyżej 60 osób, z czego dla 48 osób sieci zawsze dawały poprawne wyniki. Dla kolejnych 54 osób procent błędu nie przekraczał 50% wszystkich przejść. W grupie tej znalazły się wszystkie osoby, które nagrywane były za pomocą oprogramowania Vicon Blade.

W przypadku nagrań pochodzących z oprogramowania Vicon Nexus, znacznie częściej dochodziło do pomyłek. Wpływ na to może mieć fakt, że oprogramowanie to wykorzystywane jest do celów medycznych, w związku, z czym markery są przyklejane bezpośrednio do skóry pacjenta. Umożliwia to znacznie dokładniejszą rejestrację danych, poprzez eliminację błędów powstałych w skutek ruchu materiału. Dodatkowo, osoby zdrowe nagrywane w tym oprogramowaniu miały służyć za grupę referencyjną dla osób chorych. Ponieważ w grupie osób chorych znajdują się przede wszystkim osoby starsze, starano się, aby osoby zdrowe również były starsze.



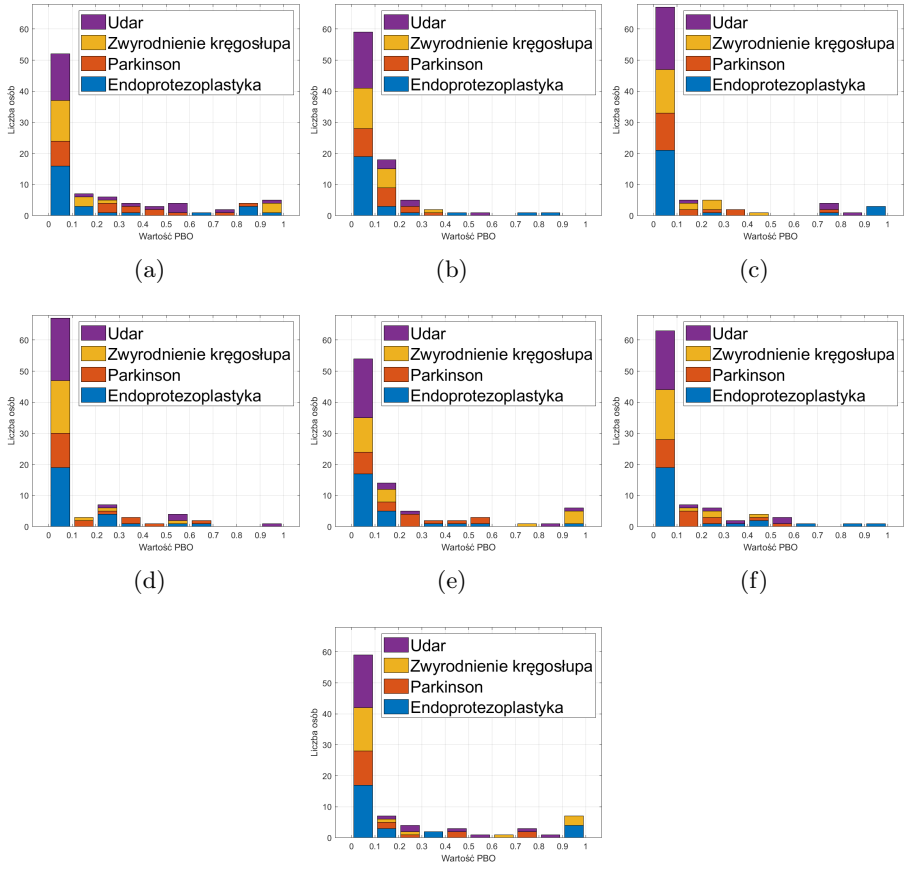
Rysunek 37: Rozkład PBO dla akcji Chód osób zdrowych dla różnej liczby znaczników w wektorze wejściowym: (a) 38 znaczników, (b) 28 znaczników, (c) 22 znaczniki, (d) 16 znaczników, (e) 13 znaczników, (f) 9 znaczników, (g) 6 znaczników.

Co warte zaznaczenia znalazły się 3 osoby, dla których sieć, niezależnie od wielkości wektora wejściowego, zawsze dawała błędne wyniki (klasyfikowała je, jako osoby chore. Dokładniejsza analiza danych tych pacjentów, nie wykazała żadnych cech wspólnych - były to osoby młode (20/21 lat), zarówno kobiety jak i mężczyzna, niezgłaszające żadnych problemów ze zdrowiem.

5.3.3 Akcja chód osób chorych

Chód osób chorych, ponownie poza jednostkowymi przypadkami, mylony był z chodem osób zdrowych. Podobnie jak w przypadku wcześniejszej grupy - nie wszystkie przejścia danej osoby były błędnie klasyfikowane, co widoczne jest na rysunku 38, gdzie przedstawiono rozkład miary PBO dla poszczególnych rozmiarów wektora wejściowego z uwzględnieniem danej jednostki chorobowej. Podobnie jak w przypadku osób zdrowych, procentowa liczba osób zawsze prawidłowo

klasyfikowanych nie spadała poniżej 50 osób, z czego dla 39 osób wszystkie sieci, zawsze dawały prawidłowe wyniki. Dodatkowo, można zauważyć, że wraz ze spadkiem liczby markerów na wejściu sieci stawały się bardziej zero-jedynkowe - więcej przejść danej osoby było uznawane za chód osoby chorej/zdrowej. Dodatkowo redukcja, w odróżnieniu od osób zdrowych, poprawiała, jakość klasyfikacji. Ponieważ, grupa ta zawiera osoby z różnymi jednostkami chorobowymi, w dalszej analizie podzielono tę grupę na poszczególne schorzenia.



Rysunek 38: Rozkład PBO dla akcji Chód osób chorych, z uwzględnieniem schorzeń, dla różnej liczby znaczników w wektorze wejściowym: (a) 38 znaczników, (b) 28 znaczników, (c) 22 znaczniki, (d) 16 znaczników, (e) 13 znaczników, (f) 9 znaczników, (g) 6 znaczników.

Wśród osób po endoprotezoplastyce stawu biodrowego zaobserwowano kilka ciekawych zależności. Osoby, których chód był zawsze lub prawie zawsze mylony z chodem osób zdrowych, przeszły zabieg endoprotezoplastyki obu stawów. Przy czym zdarzały się sytuacje, że dla różnych wektorów wejściowych osoba taka była różnie klasyfikowana, w skrajnych przypadkach wszystkie jej przejścia były albo poprawnie, albo błędnie klasyfikowane. Ogólnie, pacjenci z tym

schorzeniem dość rzadko byli błędnie identyfikowani. Wynika to z faktu, iż zabieg endoprotezoplastyki w znaczący sposób wpływa na chód człowieka, między innymi u tych osób nie występuje ruch przywiedzenia stawu biodrowego.

W grupie osób ze zwyrodnieniem kręgosłupa równie można zaobserwować bardziej skrajną klasyfikację. Poza jednostkowymi przypadkami, pacjent jest uznawany albo za osobę zdrową, albo zawsze za chorą. Przy różnym rozmiarze wektora wejściowego, inne osoby były błędnie klasyfikowane. Zdarzało się, że dana osoba była zawsze mylona tylko dla konkretnego rozmiaru wektora wejściowego. Co warte zaznaczenia, osobą, której chody były najczęściej błędnie klasyfikowane (zawsze lub prawie zawsze, jako zdrowa), była najmłodsza osoba w tej grupie - 46 letnia kobieta. Dodatkowo dla wektorów wejściowych o rozmiarze 9, 16, 22 i 28 błędnej klasyfikacji ulegało nie więcej jak 40-60% przejść danej osoby.

U osób z chorobą Parkinsona dochodziło do najmniejszej liczby pomyłek. Dla żadnego rozmiaru wektora wejściowego, nie było przypadku by dana osoba była zawsze błędnie klasyfikowana. Błędy dotyczyły pojedynczych, tych samych osób, w różnym stopniu (jedna sieć sklasyfikowała 40% przejść danej osoby, jako chód osób zdrowych inna 60%). Ponownie w grupie tej, najczęściej mylonym pacjentem okazał się najmłodszy, 34-letni mężczyzna. Dodatkowo, do pomyłek dochodziło znacznie częściej w przypadku osób, które uzyskały niższy wynik w skali UPDRS (im wyższy wynik, tym bardziej nasilone objawy choroby).

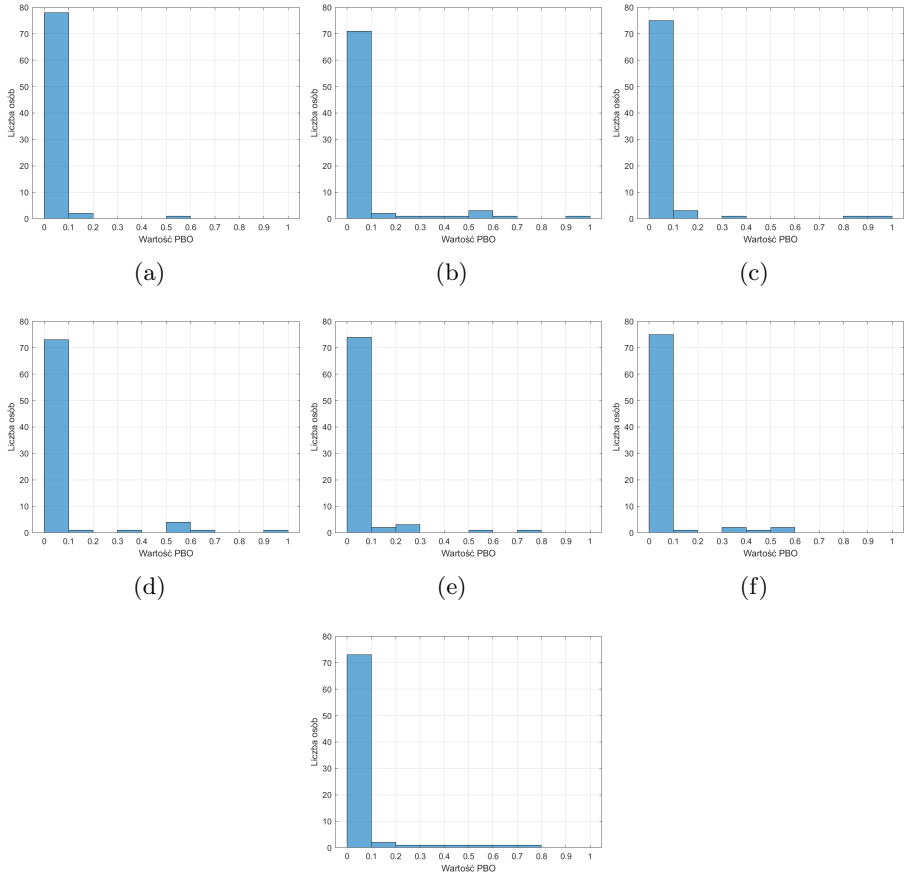
W przypadku osób po udarze, ponownie można zaobserwować bardzo skrajną klasyfikację. Ponadto, w grupie tej, przy niektórych rozmiarach wektora wejściowego (9 i 28) maksymalnie 60% przejść jednej osoby było błędnie klasyfikowanych. W pozostałych przypadkach pomyłki dotyczyły pojedynczych osób. Prawdopodobnie wynika to z faktu, iż udar w znaczącym stopniu wpływa nie tylko na chód człowieka, ale i całą jego motorykę. Tak samo jak w poprzednich grupach, najczęściej myloną osobą, niezależnie od wielkości wektora wejściowego był najmłodszy 26-letni mężczyzna z niedowładem po lewej stronie ciała.

Ogólnie wszystkie sieci znacznie lepiej poradziły sobie z prawidłową klasyfikacją osób chorych. Im bardziej dane schorzenie wpływało na chód, tym mniej przejść danej osoby było mylonych. Dodatkowo, u osoby, które były nagrywane więcej niż raz również częściej dochodziło do pomyłek. Osoby te pomiędzy nagraniami przechodziły dodatkowo rehabilitację. W przypadku, gdy dana osoba przychodziła tylko z tym jednym schorzeniem, nagrania po rehabilitacji były częściej mylone. Osoby, które miały schorzeń więcej, przykładowo były po endoprotezoplastyce stawu biodrowego, ale dodatkowo skarżyły się na bóle kręgosłupa lub cierpiały na inne schorzenia klasyfikowane były poprawnie w każdej serii nagrań. Wynika to z faktu, iż rehabilitacja skupiała się głównie na jednej dolegliwości. Niezależnie od schorzenia, jeżeli czas pomiędzy diagnozą/operacją a nagraniem był stosunkowo krótki, osoba ta była prawidłowo klasyfikowana. W sytuacji, gdy czas ten był dłuższy, a pacjent nie skarżył się na dodatkowe schorzenia, sieci klasyfikowały go w większym lub mniejszym stopniu błędnie.

5.3.4 Akcja stanie

Akcja stanie niezależnie od rozmiaru wektora wejściowego charakteryzowała się najmniejszą liczbą błędów klasyfikacji. Błędy były pojedyncze i dotyczyły praktycznie zawsze tych samych osób (rys. 39). Wśród nich znaleźli się dwaj pacjenci z chorobą Parkinsona. W ich przypadku do błędów dochodziło tylko w

przypadku większej liczby markerów. W grupie tej znalazły się też trzy osoby, dla których sieć zawsze lub prawie zawsze dawała błędne rezultaty - klasyfikowała ich ruch, jako obroty, uderzenie, czy kopnięcie.



(a) (b) (c)

(d) (e) (f)

(g)

Rysunek 39: Rozkład PBO dla akcji Stanie dla różnej liczby znaczników w wektorze wejściowym: (a) 38 znaczników, (b) 28 znaczników, (c) 22 znaczniki, (d) 16 znaczników, (e) 13 znaczników, (f) 9 znaczników, (g) 6 znaczników.

Błędy te mają związek ze sposobem wykonywania tej akcji. Zdecydowana większość osób stała dość statycznie - nie rozglądała się za mocno, najczęściej nie odrywała nóg od podłoża. Wymienione trzy osoby stały dość dynamicznie - przestępowały z nogi na nogę, wymachiwały rękami czy gwałtownie się rozglądały.

5.3.5 Akcja obroty

Pomimo stosunkowo niewielkiej reprezentacji, akcja ta okazała się na tyle specyficzna, że sieci bez problemu, niemal zawsze, prawidłowo ją klasyfikowały

(rys. 40). Podobnie jak w przypadku akcji stanie, do pomyłek dochodziło sporadycznie, zawsze w przypadku tych samych osób. Pojedyncze wykonania tej akcji przez daną osobę były mylone z akcją uderzenie, gdy wykonywała ona dynamiczne ruchy rękami, lub z akcją stanie, gdy obrót był bardzo wolny. Spadek liczby markerów w wektorze wejściowym w nieznaczny sposób pogarszał, jakość klasyfikacji.

5.3.6 Akcja schylanie się

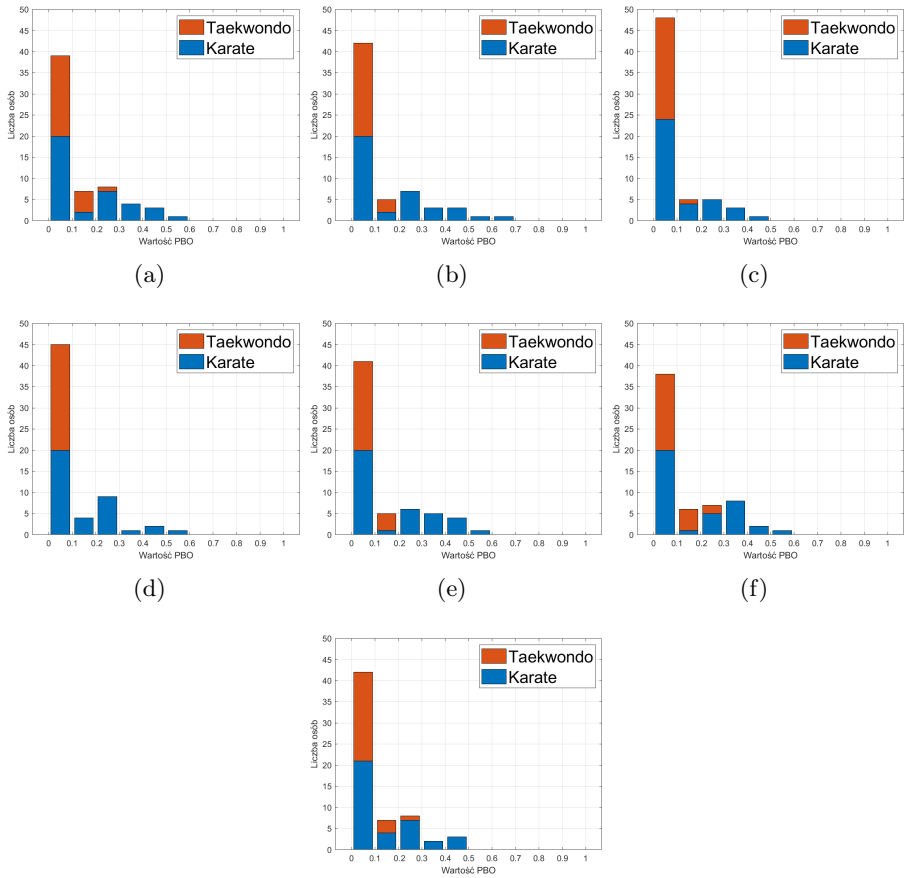
Akcja ta, podobnie jak poprzednie dwie, charakteryzowała się niemal 100% rozpoznawalnością. Do pomyłek dochodziło bardzo sporadycznie, w jednym z dwóch przypadków. Gdy skłon był zbyt niski, akcja ta mylona była ze staniem. Dodatkowe, dynamiczne ruchy rąk sprawiały, że dochodziło do pomyłki z akcją uderzenie. Akcję schylanie się wykonała największa liczba osób, z czego zdecydowana większość wykonała ją tylko raz, co wpłynęło na rozkład wartości PBO (rys. 41).

5.3.7 Akcja uderzenie

Akcje uderzenie wykonywali zawodnicy dwóch różnych sportów walki - Karate i Taekwondo. Dlatego też ponownie jak w przypadku osób chorych, na wykresie rozkładu wartości PBO oznaczono odpowiednim kolorem osoby trenujące daną dyscyplinę (rys. 42).

W przypadku osób trenujących Taekwondo, znacznie rzadziej dochodziło do pomyłek. A jeśli już pomyłki się zdarzały dotyczyły one najczęściej nie więcej niż 4-5 na 30 uderzeń danej osoby. Akcja ta najczęściej mylona była z pozostałymi akcjami niebezpiecznymi - kopnięcia. Do pomyłek dochodziło w momencie, gdy dana osoba podczas wykonywania uderzenia dodatkowo przemieszczała się. Miało to miejsce najczęściej podczas uderzenia w tarczę lub deskę. W jednostkowych przypadkach akcja ta mylona była też z akcją stanie lub obroty. W grupie tej, najlepsze rezultaty osiągnięto przy wektorze wejściowym składającym się z 16 znaczników, jednakże błędy dla pozostałych wartości nie były specjalnie większe.

Wśród zawodników karate znacznie częściej dochodziło do pomyłek. Tak samo jak w przypadku wszystkich pozostałych grup, pomyłki te dotyczyły najczęściej tych samych osób. Podobnie jak w przypadku zawodników Taekwondo, pomyłki dotyczyły głównie uderzenia w tarczę - uderzenia w powietrze danej osoby były poprawnie sklasyfikowane, a w tarcze mylone z innymi akcjami niebezpiecznymi. Prawdopodobna przyczyna takiej klasyfikacji może mieć podłoże w tym, że podczas uderzenia w tarczę, zawodnicy wykonywali kilka dodatkowych ruchów.

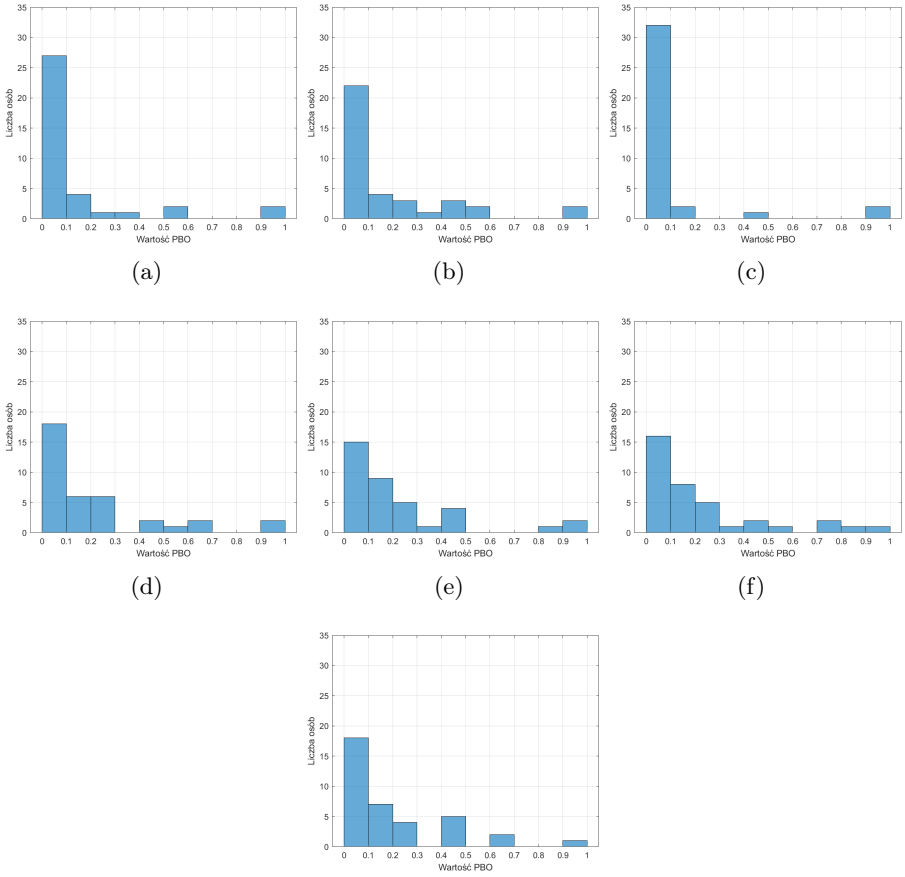


Rysunek 42: Rozkład PBO dla akcji Uderzenie, z uwzględnieniem trenowanego sportu, dla różnej liczby znaczników w wektorze wejściowym: (a) 38 znaczników, (b) 28 znaczników, (c) 22 znaczniki, (d) 16 znaczników, (e) 13 znaczników, (f) 9 znaczników, (g) 6 znaczników.

5.3.8 Akcja kopnięcie niskie

Akcję tą wykonywali tylko zawodnicy karate, a jej rozpoznawalność w bardzo dużej mierze zależy od rozmiarów wektora wejściowego (rys. 43). Spośród wszystkich akcji niebezpiecznych, akcja ta charakteryzowała się największą ilością pomyłek. Początkowo założono, że wpływ na tak częste pomyłki może mieć fakt, iż ten rodzaj kopnięcia wykonywali tylko zawodnicy karate, wśród, których dominowały osoby młodsze (od 10 do 14 lat). Dodatkowo przemawiał za tym fakt, że dwie osoby, dla których wszystkie sieci zawsze dawały błędne wyniki, były dorosłe (20 i 50 lat) i wysokie (ponad 180 cm wzrostu). Jednakże dla pozostały osób pełnoletnich, lub wyższych nieletnich rezultaty były znacznie lepsze, lub nawet za każdym razem bezbłędne. Dalsza analiza nagrań wykazała, że znacznie częściej do pomyłek dochodziło przy kopnięciach w powietrze. Brak celu w postaci tarczy powodował, że wysokością niemalże zrównywały się

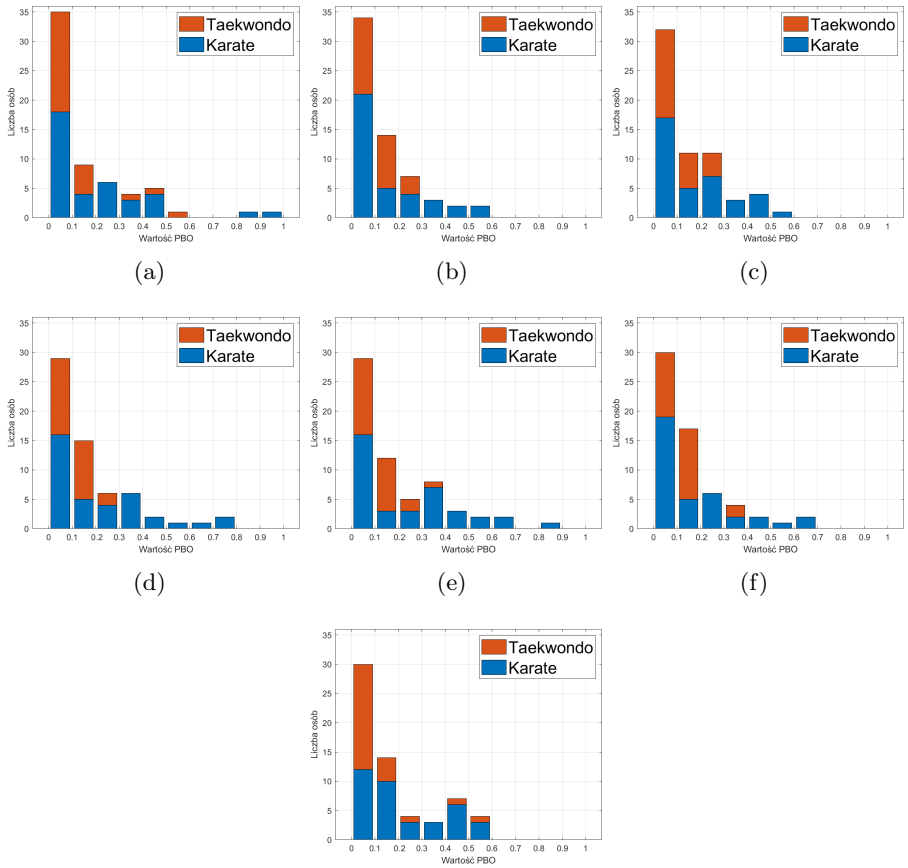
z kopnięciami wysokimi.



Rysunek 43: Rozkład PBO dla akcji Kop niski dla różnej liczby znaczników w wektorze wejściowym: (a) 38 znaczników, (b) 28 znaczników, (c) 22 znaczniki, (d) 16 znaczników, (e) 13 znaczników, (f) 9 znaczników, (g) 6 znaczników.

5.3.9 Akcja kopnięcie wysokie proste

Akcje kopnięcie wysokie proste, ponownie wykonywali zawodnicy zarówno Karate jak i Taekwondo. Rozkład wartości PBO jest podobny jak przy akcji uderzenie - dane zawodników Taekwondo są częściej prawidłowo klasyfikowane niż zawodników Karate (rys. 42). Ponownie częściej mylone są kopnięcia wykonane w tarczę lub w deskę, podczas których zawodnik wykonuje dodatkowe ruchy. Prowadziło to do pomyłek z pozostałymi akcjami niebezpiecznymi. Pomyłki z akcją kopnięcie niskie, zdarzały się, gdy tarcza/deska znajdowała się stosunkowo nisko. Z akcją tą myleni byli zarówno zawodnicy Karate jak i Taekwondo.



(a) (b) (c)

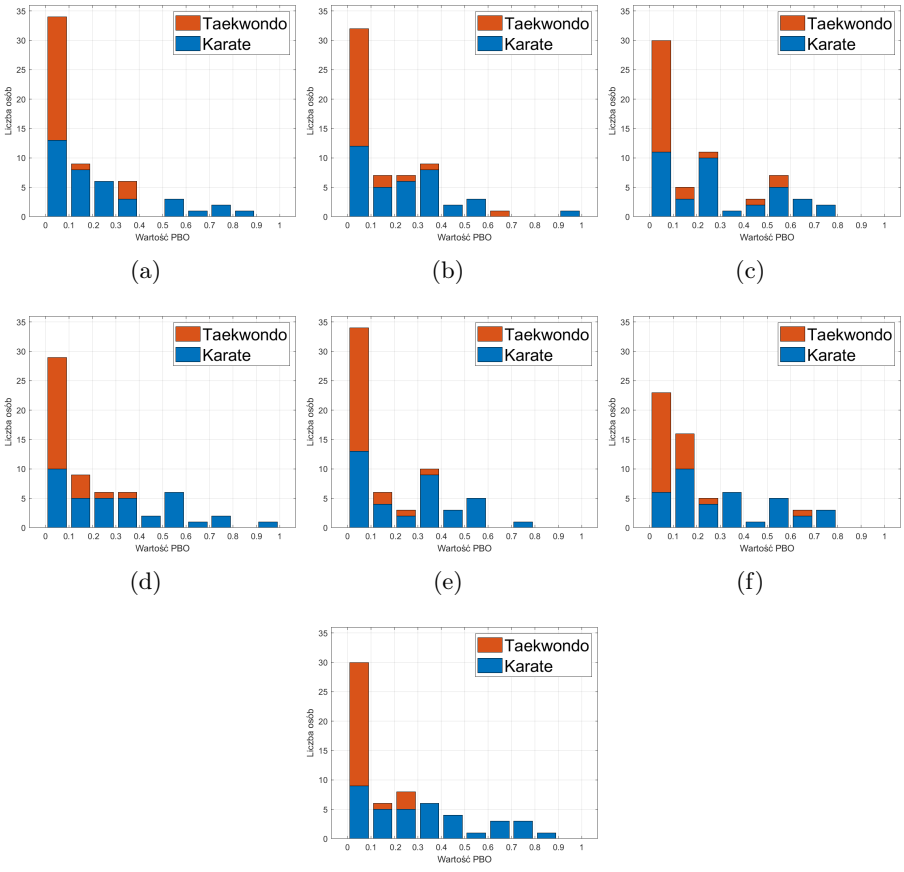
(d) (e) (f)

(g)

Rysunek 44: Rozkład PBO dla akcji Kopnięcie wysokie proste, z uwzględnieniem trenowanego sportu, dla różnej liczby znaczników w wektorze wejściowym: (a) 38 znaczników, (b) 28 znaczników, (c) 22 znaczniki, (d) 16 znaczników, (e) 13 znaczników, (f) 9 znaczników, (g) 6 znaczników.

5.3.10 Akcja kopnięcie wysokie boczne

Ostatnią z omawianych akcji jest kopnięcie wysokie boczne, wykonywane przez zawodników obu dyscyplin. Charakterystyka tej akcji jest podobna do pozostałych akcji niebezpiecznych (rys. 45) - kopnięcia wykonane przez zawodników Taekwondo były częściej poprawnie klasyfikowane, niż zawodników Karate. Przyczyna błędnej klasyfikacji również leży w sposobie wykonania akcji - częściej mylone były kopnięcia wykonywane w tarcze lub deskę.



(a) (b) (c)

(d) (e) (f)

(g)

Rysunek 45: Rozkład PBO dla akcji Kopnięcie wysokie boczne, z uwzględnieniem trenowanego sportu, dla różnej liczby znaczników w wektorze wejściowym: (a) 38 znaczników, (b) 28 znaczników, (c) 22 znaczniki, (d) 16 znaczników, (e) 13 znaczników, (f) 9 znaczników, (g) 6 znaczników.

5.3.11 Wnioski

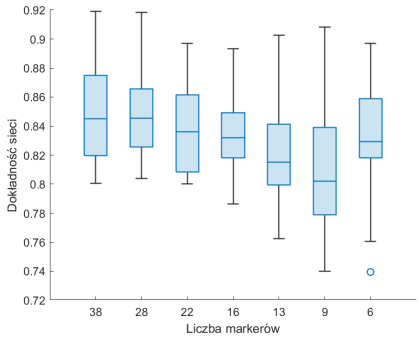
Dwukierunkowe sieci LSTM dość dobrze poradziły sobie z klasyfikacją wybranych akcji. Do pomyłek dochodziło najczęściej pomiędzy bardzo podobnymi akcjami. Większość pomyłek miała miejsce w obrębie bardzo podobnych klas - chody, czy różne rodzaje kopnięć. W przypadku obu akcji chód możemy zauważyć, że większość sieci była mocno wyczulona na wszelkie odstępstwa od pewnego wzorca. Powodowało to, że osoby zdrowe znacznie częściej klasyfikowane były, jako chore. W przypadku akcji potencjalnie niebezpiecznych największe wpływ na poprawną rozpoznawalność miał sposób wykonania techniki - kopnięcie/uderzenie w powietrze lub tarczę/deskę.

Rozmiar wektora wejściowego, w przypadku większości akcji, miał bardzo duży wpływ, na jakość klasyfikacji. Szczegółowa analiza danych wykazała, że najbardziej optymalny jest wektor wejściowy składający się z 22 markerów.

5.4 Klasyfikacja z wykorzystaniem sieci CNN

5.4.1 Ogólna, jakość klasyfikacji

Sieci CNN, niezależnie od rozmiarów wektora wejściowego, uzyskały gorsze rezultaty w stosunku do dwukierunkowej sieci LSTM, jednakże lepsze niż jednokierunkowy LSTM. Co ważniejsze, w odróżnieniu od jednokierunkowej sieci LSTM dochodziło do większej liczby pomyłek w ramach podobnych akcji, a nie pomiędzy klasami, co widoczne jest na rysunku 48. Dodatkowo, nie miała miejsca sytuacja by wszystkie warianty danej akcji klasyfikowane były, jako jedna akcja (przykładowo wszystkie rodzaje kopnięć, jako kopnięcie wysokie). Dlatego też, ten rodzaj sieci został poddany dalszej analizie.



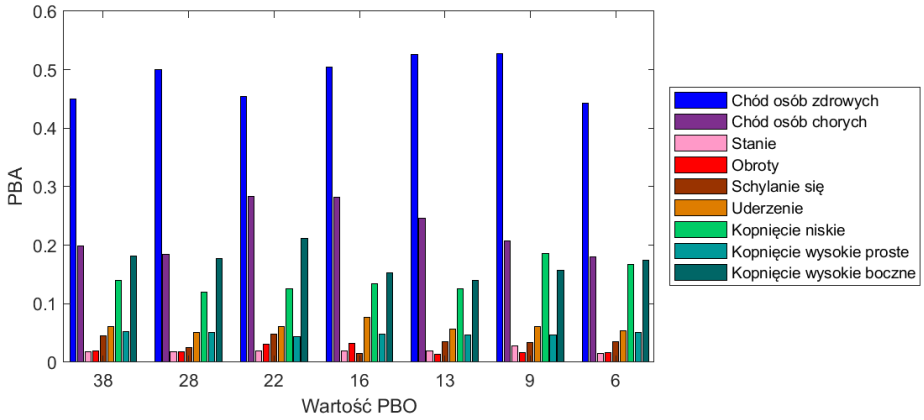
Rysunek 46: Wykres pudełkowy dla sieci CNN

Zależność pomiędzy dokładnością sieci CNN a rozmiarem wektora wejściowego jest inna niż w przypadku sieci LSTM. Dokładność systematycznie spada wraz ze spadkiem liczby markerów na wejściu (rys. 46). Natomiast wahania pomiędzy dokładnościami poszczególnych sieci, podobnie jak w przypadku sieci LSTM, rosły wraz ze zmniejszaniem się wektora wejściowego i wynosiły od 9,5% do 17%. Dla 38, 28, 22 i 16 markerów na wejściu spadek średniej, jakości jest nieznaczny. Co więcej dla wektora wejściowego składającego się z 22 znaczników, wahania dokładności są najmniejsze. W przypadku 6 markerów na wejściu

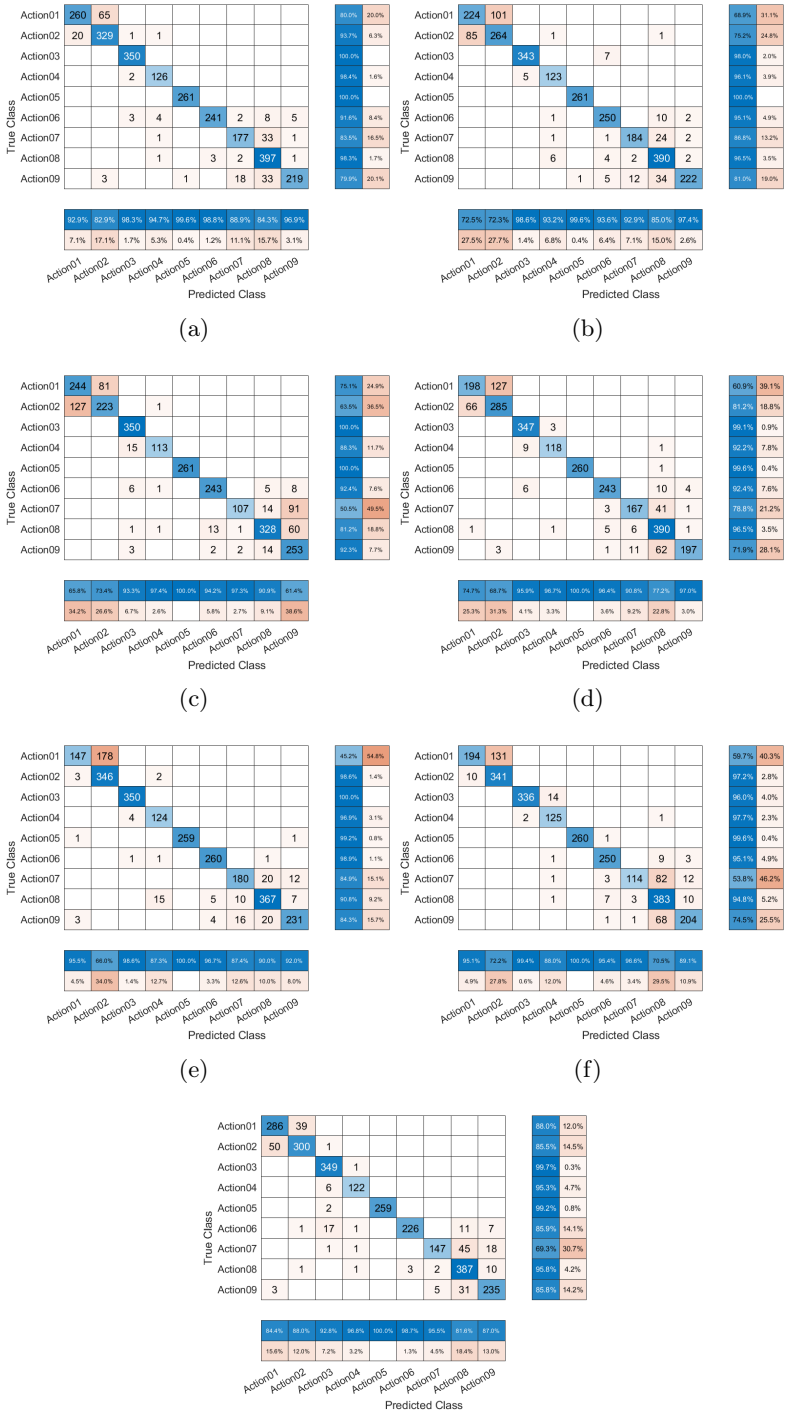
można zaobserwować znaczący wzrost średniej, jednakże rozbieżność pomiędzy najlepszą a najgorszą siecią była bardzo duża.

Jak już wcześniej wspomniano, do zdecydowanej większości pomyłek dochodziło pomiędzy bardzo podobnymi akcjami (chód osób zdrowych i chód osób chorych, czy różne rodzaje kopnięć). Na rysunku 48 przedstawiono macierze pomyłek dla najlepszych sieci dla każdego rozmiaru wektora wejściowego. Podobnie jak w przypadku sieci LSTM, akcje statyczne charakteryzowały się nie tylko największą rozpoznawalnością, ale też rzadko inne były z nimi mylone. Ponownie wraz ze spadkiem liczby markerów w wektorze wejściowym można zauważyć zwiększenie się pomyłek między klasowych.

Największy udział w jakości ponownie miała akcja chód osób zdrowych i chód osób chorych. Jest to szczególnie widoczne na średnim procentowym rozkładzie błędów dla poszczególnych akcji (rys. 47). Niezależnie od rozmiarów wektora wejściowego około połowa osób zdrowych była błędnie klasyfikowana. Wszystkie wyniki zostały ponownie poddane bardziej szczegółowej analizie, osobno dla każdej akcji.



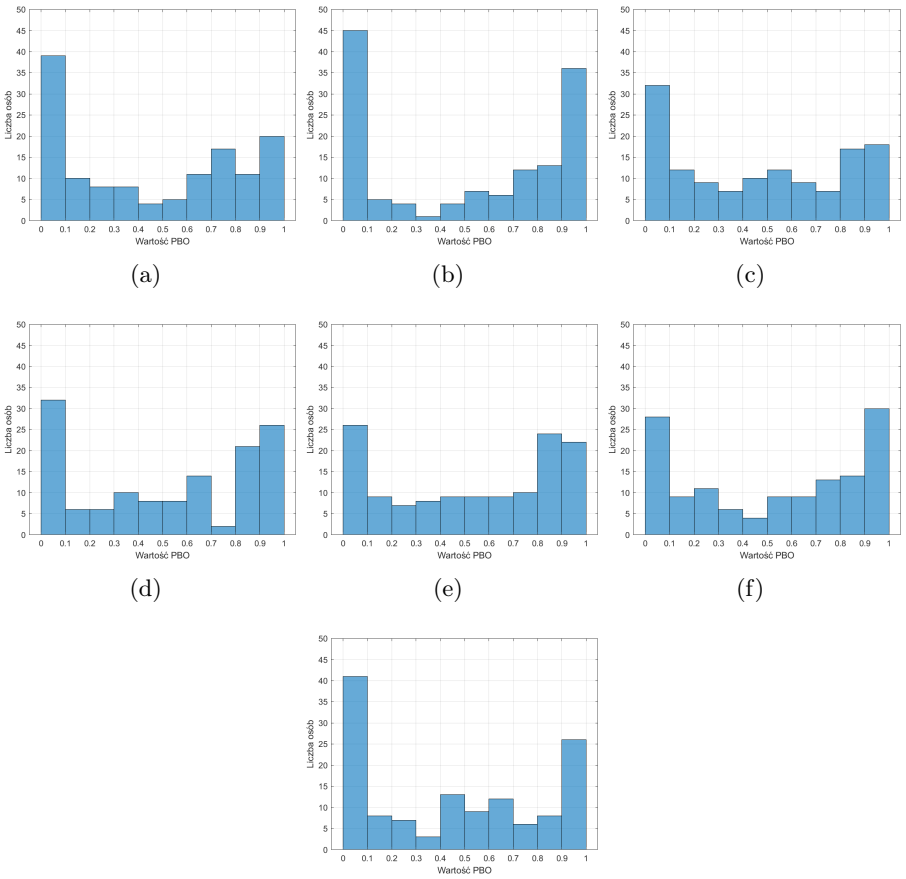
Rysunek 47: Wykresy słupkowe wartości PBA dla różnej liczby markerów w wektorze wejściowym



Rysunek 48: Macierze pomyłek sieci CNN dla różnej liczby znaczników w wektorze wejściowym: (a) 38, (b) 28, (c) 22, (d) 16, (e) 13, (f) 9, (g) 7

5.4.2 Akcja Chód osób zdrowych

Jak już wspomniano wcześniej, akcja ta była najczęściej myloną akcją wśród wszystkich utworzonych sieci. Na rysunku 49 przedstawiono średni rozkład wartości PBO. Odzwierciedla on ogólną, jakość sieci - wraz ze spadkiem liczby markerów w wektorze wejściowym, liczba osób poprawnie klasyfikowanych spada, by na końcu dla 6 znaczników ponownie wzrosnąć.



(a) (b) (c)

(d) (e) (f)

(g)

Rysunek 49: Rozkład PBO dla akcji Chód osób zdrowych dla różnej liczby znaczników w wektorze wejściowym: (a) 38 znaczników, (b) 28 znaczników, (c) 22 znaczniki, (d) 16 znaczników, (e) 13 znaczników, (f) 9 znaczników, (g) 6 znaczników.

Dokładniejsza analiza danych wykazała, że ponownie osoby nagrywane za pomocą oprogramowania Vicon Blade były częściej poprawnie klasyfikowane. Ruch wszystkich 27 osób, które niezależnie od rozmiaru wektora wejściowego były zawsze poprawnie klasyfikowane został zarejestrowany właśnie w tym oprogramowaniu. Przy czym zdarzały się osoby, nagrywane w Vicon Blade, których część przejść była błędnie klasyfikowana.

Znalazło się też 9 osób, których chód zawsze był mylony z chodem osoby chorej. Wśród nich były 3 osoby zawsze mylone przez sieci LSTM, oraz 6 osób, dla których sieci LSTM myliły się średnio w 50%. Dokładniejsza analiza dostępnych informacji na temat tych osób nie wykazała żadnych nieprawidłowości. Podgląd ich nagrań również nie wykazał żadnych widocznych odstępstw. Przyczyna musi, więc leżeć w bardziej abstrakcyjnych cechach znalezionych przez sieci.

5.4.3 Akcja Chód osób chorych

Akcja chód osób chorych, poza jednostkowymi przypadkami, mylona była z chodem osób zdrowych. Do sytuacji, w której wszystkie przejścia danej osoby były błędnie klasyfikowane dochodziło rzadko (rys. 50). Ponownie jak w przypadku osób zdrowych możemy zaobserwować wzrost liczby osób poprawnie klasyfikowanych wraz ze spadkiem liczby markerów w wektorze wejściowym. Dla 35 osób sieci niezależnie od wielkości wektora wejściowego zawsze dawały prawidłowe wyniki. Rozkład błędów dla poszczególnych jednostek chorobowych przestawiał się następująco.

W grupie pacjentów po endoprotezoplastyce dochodziło do sporadycznych pomyłek. Błędy najczęściej dotyczyły kilku przejść danej osoby. Tylko dla 3 rozmiarów wektora wejściowego znalazły się osoby, które sieć zawsze klasyfikowała nieprawidłowo. Osoby te przeszły zabieg endoprotezoplastyki obu stawów biodrowych. Pozostałe osoby, których chód był często mylony również miały operowane oba biodra, lub były po zabiegu jednego i oczekiwały na zabieg drugiego stawu biodrowego. Zależność ta nie jest jednak symetryczna - część osób po podwójnym zabiegu była zawsze prawidłowo klasyfikowana. Osoby te najczęściej cierpiały na inne dodatkowe schorzenia jak osteopatia, bóle kręgosłupa, czy zwyrodnienie stawów kolanowych.

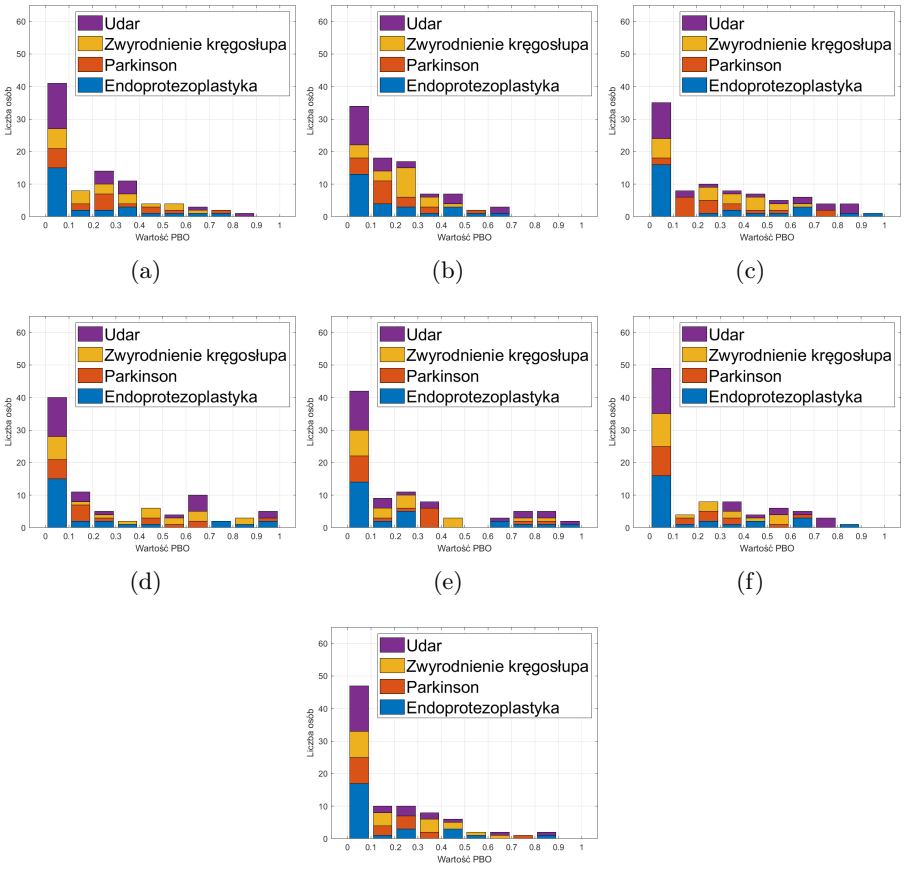
Liczba zawsze poprawnie klasyfikowanych pacjentów ze zwyrodnieniem kręgosłupa również rosła wraz ze spadkiem liczby markerów. Jednakże maksymalnie połowa była zawsze poprawnie klasyfikowana. Osoby te poza problemami z kręgosłupem zgłaszały inne dodatkowe dolegliwości jak bóle stawów biodrowych, czy osteopatię. Osobami, których chód był najczęściej mylony były osoby, które skarżyły się głównie na bóle kręgosłupa tylko w odcinku lędźwiowym. Pozostałe osoby, których chód tylko w części przypadków był mylony, nagrywane były w laboratorium więcej niż raz. Pomiędzy kolejnymi nagraniami przechodziły sesje rehabilitacji, co wpłynęło na ich sposób poruszania się.

Tak samo jak dla poprzednich jednostek chorobowych, klasyfikacja osób z chorobą Parkinsona poprawiała się wraz ze zmniejszaniem się wektora wejściowego. Przy czym dla wektora wejściowego składającego się z 22 markerów, można zauważyć drastyczny spadek, jakości - tylko 2 osoby były poprawnie identyfikowane. W przypadku sieci CNN, w odróżnieniu od LSTM, nie widać tak wyraźnej korelacji, pomiędzy jakością klasyfikacji a punktacją pacjenta w skali UPDRS.

Osoby po udarze były błędnie klasyfikowane w różnym stopniu. Ponownie osoby, które posiadały dodatkowe schorzenia (haluksy, cukrzyca, bóle kończyn dolnych) lub u których czas pomiędzy udarem a nagraniem był krótki były zawsze poprawnie klasyfikowane. Przy czym zdarzały się wyjątki. Przykładowo pacjent, u którego udar wystąpił w tym samym roku, co nagranie był dość często mylony. Możliwych przyczyn takiej klasyfikacji może być wiele. Należy wziąć pod uwagę, że udar w zależności od rozmiarów i położenia upośledza w

różnym stopniu funkcje motoryczne.

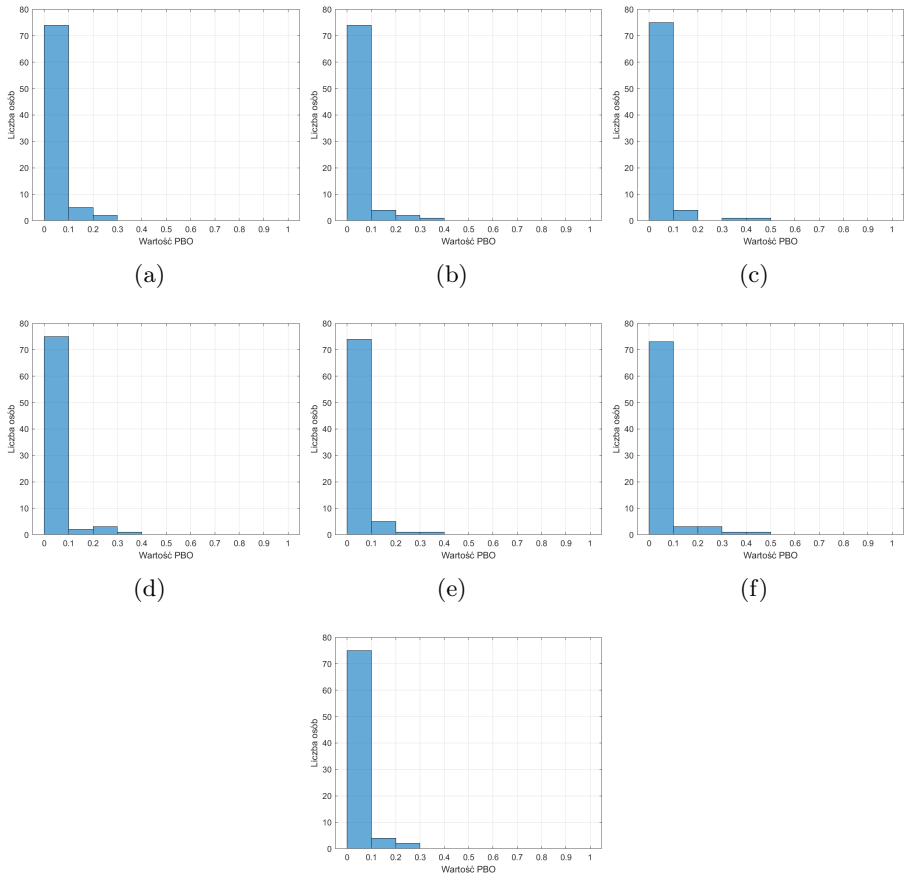
Wśród wszystkich pacjentów miała miejsce podobna zależność jak przy chodzie osób zdrowych. Wartość PBO uzyskana dla sieci CNN była dla każdej osoby podobna lub wyższa niż te uzyskane dla sieci LSTM. Zdarzały się jednosetowe sytuacje odwrotne. Ponieważ błędy dotyczyły w większości tych samych osób, ich przyczyny są podobne jak przy sieciach LSTM.



Rysunek 50: Rozkład PBO dla akcji Chód osób chorych, z uwzględnieniem schorzeń, dla różnej liczby znaczników w wektorze wejściowym: (a) 38 znaczników, (b) 28 znaczników, (c) 22 znaczniki, (d) 16 znaczników, (e) 13 znaczników, (f) 9 znaczników, (g) 6 znaczników.

5.4.4 Akcja Stanie

Akcja stanie, niezależnie od rozmiarów wektora wejściowego, była praktycznie zawsze poprawnie klasyfikowana. W przypadku tej akcji sieci CNN poradziły sobie znacznie lepiej niż sieci LSTM - błędy dotyczyły maksymalnie połowy nagrań danej osoby (rys. 51). Prawdopodobna przyczyna sporadycznych pomyłek, leży, podobnie jak w przypadku sieci LSTM, w sposobie wykonywania danej akcji przez danę osobę. Nadmierne ruchy kończyn, oraz gwałtowne obroty czy przestępowanie z nogi na nogę, powodowały błędną klasyfikację jako uderzenie, obroty czy chód.



(a) (b) (c)

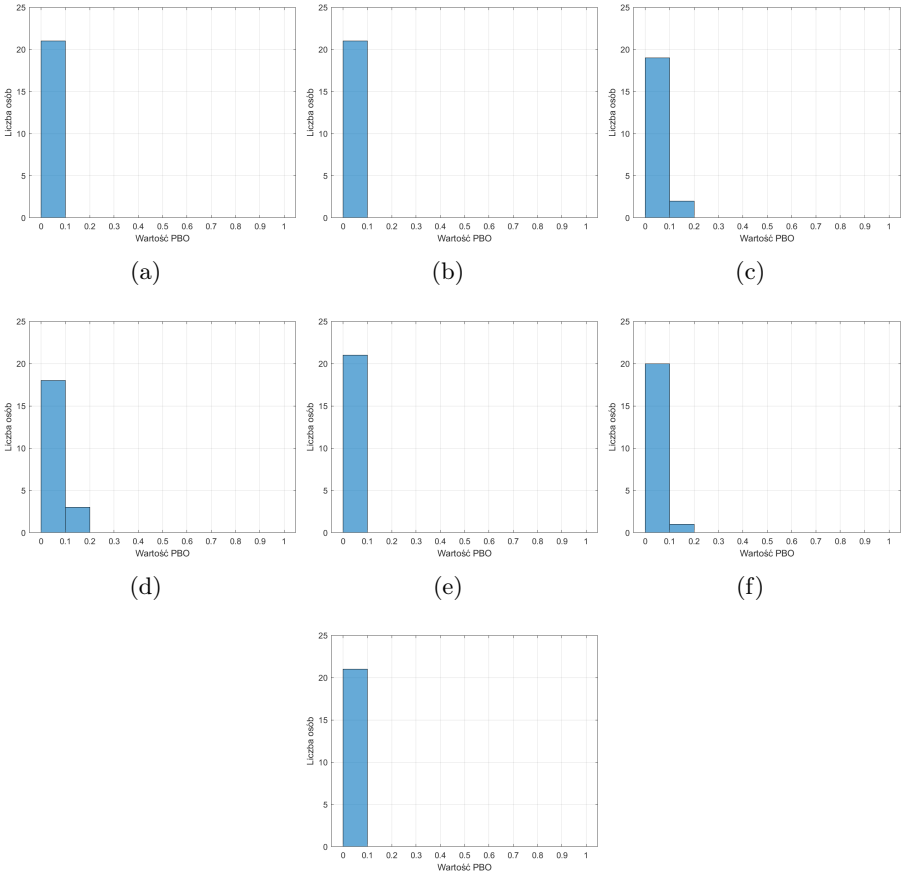
(d) (e) (f)

(g)

Rysunek 51: Rozkład PBO dla akcji Stanie dla różnej liczby znaczników w wektorze wejściowym: (a) 38 znaczników, (b) 28 znaczników, (c) 22 znaczniki, (d) 16 znaczników, (e) 13 znaczników, (f) 9 znaczników, (g) 6 znaczników.

5.4.5 Akcja Obroty

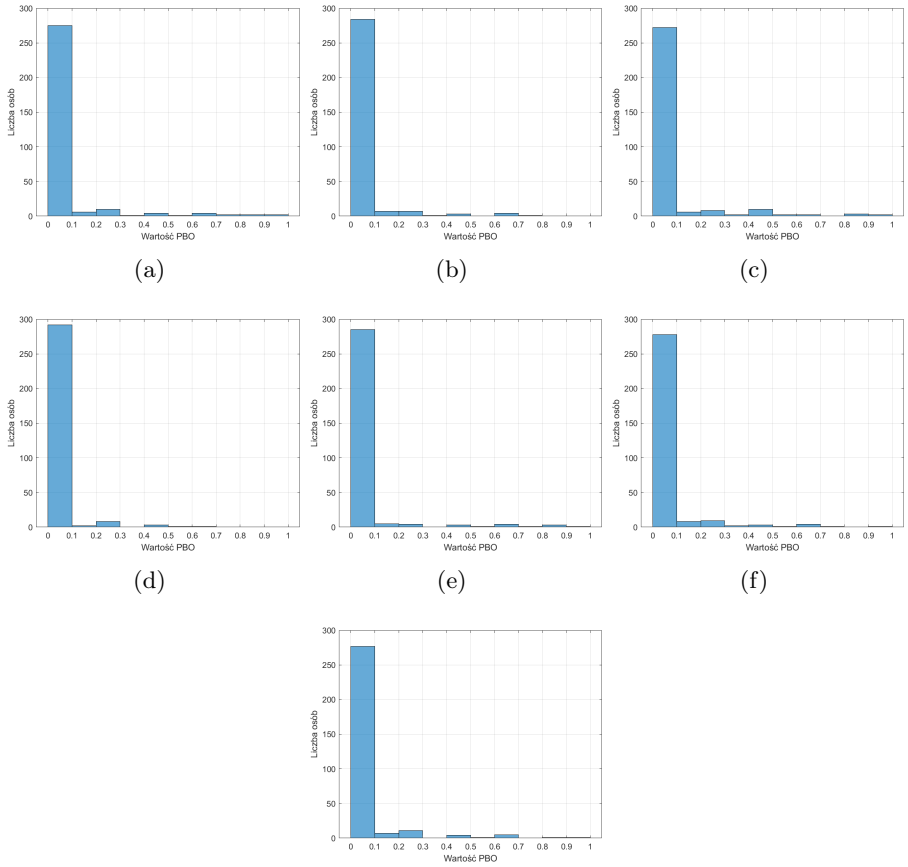
Akcja obroty, niezależnie od wielkości wektora wejściowego, była niemal w 100% prawidłowo klasyfikowana (rys. 52). Wyjątek stanowiły dwie osoby, które podczas niektórych nagrań bardziej dynamicznie wymachiwali rękami (klasyfikowane, jako uderzenie), lub obracali się bardzo powoli (klasyfikowane, jako stanie).



Rysunek 52: Rozkład PBO dla akcji Obrót dla różnej liczby znaczników w wektorze wejściowym: (a) 38 znaczników, (b) 28 znaczników, (c) 22 znaczniki, (d) 16 znaczników, (e) 13 znaczników, (f) 9 znaczników, (g) 6 znaczników.

5.4.6 Akcja Schylanie się

Podobnie jak dwie poprzednie, bardziej statyczne akcje, akcja schylanie się była w większości przypadków prawidłowo klasyfikowana. Do pomyłek dochodziło bardzo sporadycznie - przy bardzo niskim skłonie, lub jakiś dodatkowych ruchach kończyn górnych. Na rysunku 53 przedstawiono rozkład wartości PBO dla poszczególnych rozmiarów wektora wejściowego. Zmiana markerów na wejściu nieznacznie wpłynęła na rozpoznawalność tej akcji.

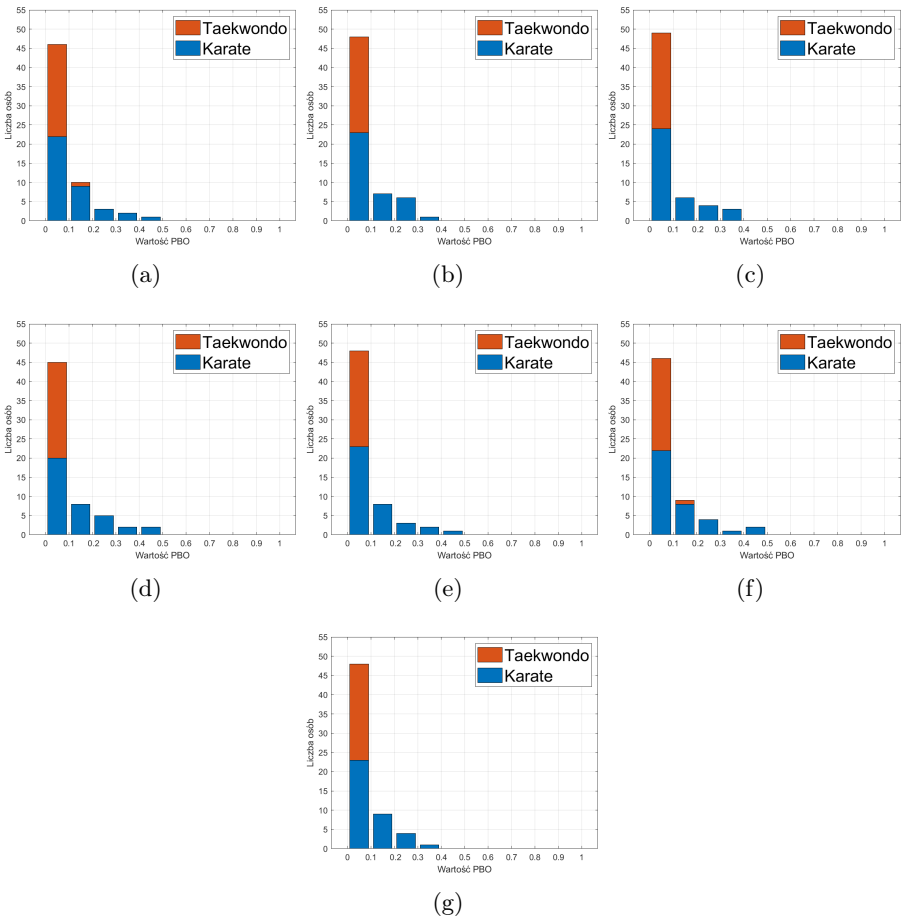


Rysunek 53: Rozkład PBO dla akcji Schylanie się dla różnej liczby znaczników w wektorze wejściowym: (a) 38 znaczników, (b) 28 znaczników, (c) 22 znaczniki, (d) 16 znaczników, (e) 13 znaczników, (f) 9 znaczników, (g) 6 znaczników.

5.4.7 Akcja Uderzenie

W przypadku pierwszej z "niebezpiecznych" akcji - uderzenie, sieć CNN poradziła sobie znacznie lepiej. Zdecydowana większość uderzeń danej osoby była klasyfikowana poprawnie (rys. 54). W przypadku zawodników trenujących Taekwondo do pomyłek praktycznie nie dochodziło. Wyjątek stanowiła jedna osoba, dla której przy 38 i 9 markerach na wejściu, sieć niepoprawnie sklasyfikowała 3 uderzenia w tarczę i 2 uderzenia w deskę. Akcje te klasyfikowane były, jako stanie, lub kopnięcie niskie.

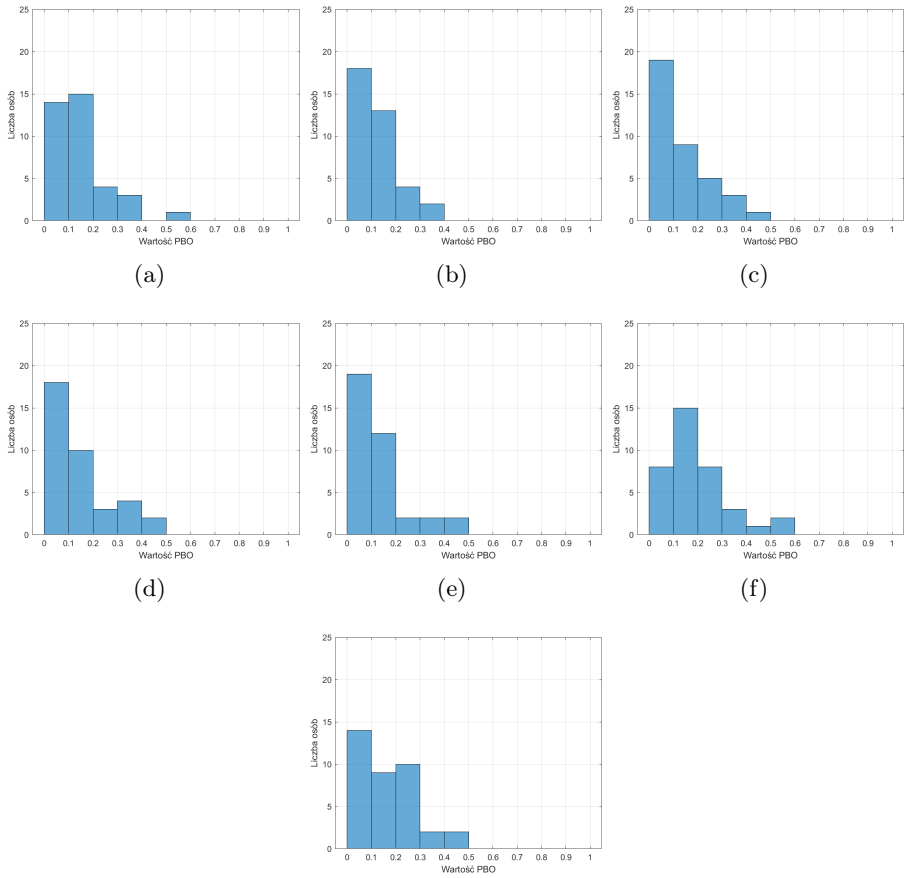
Wśród zawodników Karate, znacznie częściej dochodziło do błędów. Pomyłki, niezależnie od wielkości wektora wejściowego dotyczyły tych samych osób - dla 38 markerów na wejściu sieć błędnie klasyfikowała 40% uderzeń osoby, a dla 16 25% uderzeń. Do pomyłek dochodziło głównie, gdy zawodnik uderzał w tarczę.



Rysunek 54: Rozkład PBO dla akcji Uderzenie, z uwzględnieniem trenowanego sportu, dla różnej liczby znaczników w wektorze wejściowym: (a) 38 znaczników, (b) 28 znaczników, (c) 22 znaczniki, (d) 16 znaczników, (e) 13 znaczników, (f) 9 znaczników, (g) 6 znaczników.

5.4.8 Akcja Kopnięcie niskie

Akcja kopnięcie niskie, wykonywana tylko przez zawodników Karate była znacznie gorzej klasyfikowana niż pozostałe akcje "niebezpieczne". Niezależnie od rozmiarów wektora wejściowego, ponownie nie więcej jak połowa kopnięć danej osoby było błędnie klasyfikowanych (rys. 55). Analiza konkretnych nagrań wykazała, iż podobnie jak w przypadku sieci LSTM do pomyłek częściej dochodziło w przypadku kopnięć w powietrze. Brak tarczy powodował, że zawodnik unosił nogę wyżej, przez co sieć klasyfikowała ten rodzaj kopnięcia, jako kopnięcie wysokie proste lub boczne.



(a) (b) (c)

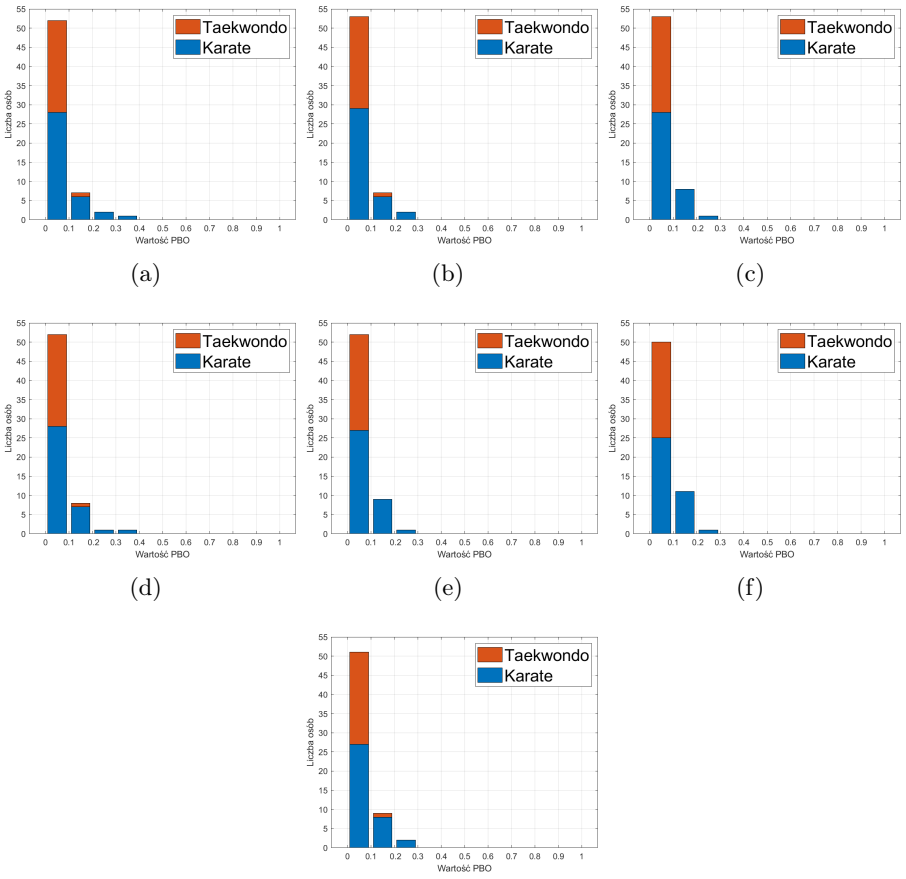
(d) (e) (f)

(g)

Rysunek 55: Rozkład PBO dla akcji Kop niski dla różnej liczby znaczników w wektorze wejściowym: (a) 38 znaczników, (b) 28 znaczników, (c) 22 znaczniki, (d) 16 znaczników, (e) 13 znaczników, (f) 9 znaczników, (g) 6 znaczników.

5.4.9 Akcja kopnięcie wysokie proste

Ten rodzaj kopnięcia wykonywany był przez zawodników obu sportów walki. Rozkład wartości PBO jest podobny jak przy akcji uderzenie (rys. 54). Rozmiar wektora wejściowego nie ma większego wpływu, na jakość klasyfikacji, a błędy dotyczą głównie tych samych zawodników Karate. Wśród zawodników Taekwondo dochodzi do sporadycznych pomyłek podczas uderzenia w tarczę lub deskę. Od wysokości i ustawienia celu zależało czy ruch ten będzie pomylony z akcją kopnięcie niskie, czy z kopnięciem wysokim bocznym.



(a) (b) (c)

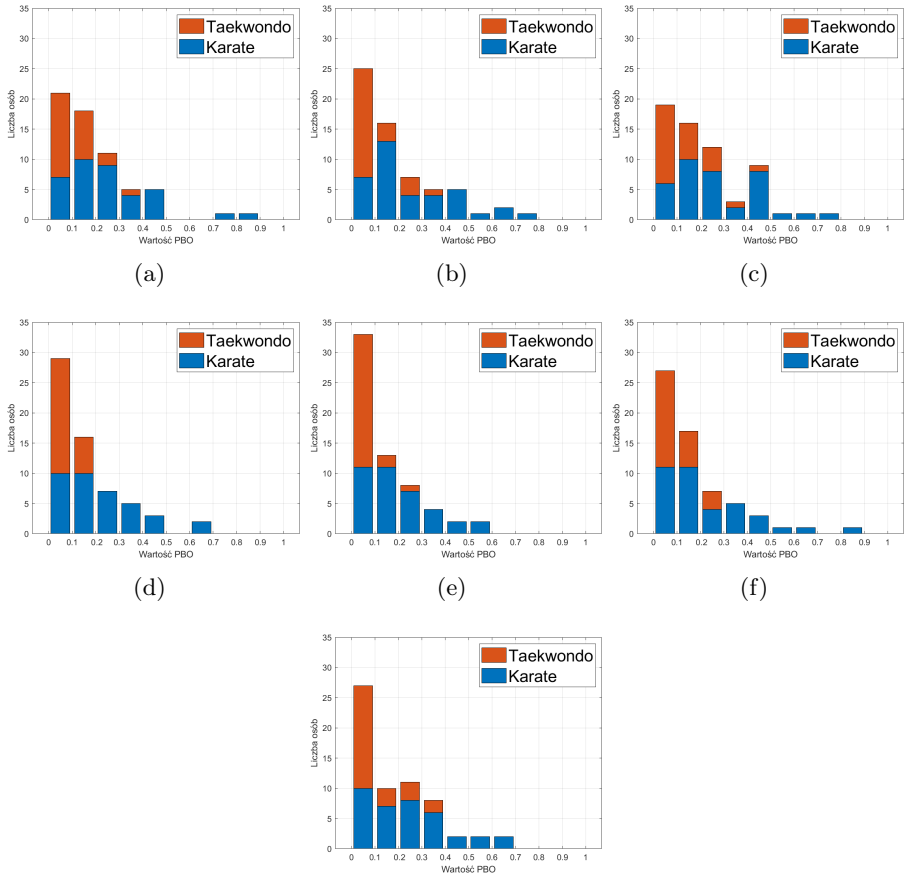
(d) (e) (f)

(g)

Rysunek 56: Rozkład PBO dla akcji Kopnięcie wysokie proste, z uwzględnieniem trenowanego sportu, dla różnej liczby znaczników w wektorze wejściowym: (a) 38 znaczników, (b) 28 znaczników, (c) 22 znaczniki, (d) 16 znaczników, (e) 13 znaczników, (f) 9 znaczników, (g) 6 znaczników.

5.4.10 Akcja kopnięcie wysokie boczne

Dla ostatniej z wybranych akcji, ponownie można zauważyć wpływ rozmiaru wektora wejściowego, na jakość klasyfikacji. Mniejsza liczba markerów na wejściu znacząco poprawia, jakość klasyfikacji (rys. 57). Większość błędów ponownie dotyczy zawodników Karate, jednakże w odróżnieniu od pozostałych akcji "niebezpiecznych", wśród zawodników Taekwondo również dochodziło do większej liczby pomyłek.



Rysunek 57: Rozkład PBO dla akcji Kopnięcie wysokie boczne, z uwzględnieniem trenowanego sportu, dla różnej liczby znaczników w wektorze wejściowym: (a) 38 znaczników, (b) 28 znaczników, (c) 22 znacznik, (d) 16 znaczników, (e) 13 znaczników, (f) 9 znaczników, (g) 6 znaczników.

5.4.11 Wnioski

Sieci CNN, pomimo niższej średniej wartości dokładności, poradziły sobie z zadaniem klasyfikacji całkiem dobrze, a dla niektórych akcji przewyższyły sieci LSTM. Największy wpływ, na jakość sieci miała ponownie akcja chód osób zdro-

wych i chód osób chorych. Liczne błędy, zwłaszcza dla pierwszej z wymienionych akcji, w znaczący sposób obniżały, jakość sieci.

Wpływ rozmiaru wektora wejściowego, na jakość klasyfikacji różnił się w zależności od akcji. W przypadku akcji statycznych nie miał on większego wpływu, dla pozostałych zazwyczaj mniejsza liczba markerów na wejściu poprawiała, jakość klasyfikacji. Po szczegółowej analizie wszystkich danych uznano, że najbardziej optymalny jest wektor wejściowy składający się z 16 markerów.

5.5 Porównanie wyników dla obu rodzajów sieci

Błędy w obu omówionych wyżej sieciach były bardzo podobne. Dotyczy to zarówno klas, które były miedzy sobą mylone, jak i osób, których ruch był mylony. Przykładowo sieć LSTM klasyfikowała średnio 3 z 10 przejść jednej osoby niepoprawnie, a sieć CNN 4 z 10. Przy czym nie w każdym przypadku były to te same nagrania.

5.5.1 Akcje - Chody

Najczęściej mylonymi akcjami były akcje chód osób zdrowych i chód osób chorych. Znacznie częściej dochodziło do uznawania osoby zdrowej za chorą niż odwrotnie. Wszystkie sieci wydawały się być bardzo rygorystyczne w klasyfikacji - osoby zdrowe, których chód odbiegał w mniejszym lub większym stopniu od wzorca klasyfikowane były, jako chore. Wśród takich osób możemy wyróżnić 24 letnią kobietę, która chodziła na palcach, lub osoby, które podczas poruszania się mocno wymachiwały rękami.

Dodatkowo do pomyłek znacznie częściej dochodziło wśród osób nagrywanych w oprogramowaniu Vicon Nexus. Wpływ na taką klasyfikacje może mieć nie tyle oprogramowanie, a specyfikacja nagrań. Osoby nagrywane w Vicon Nexus miały służyć za grupę referencyjną dla osób z różnymi schorzeniami, dlatego też ich średnia wieku jest wyższa niż dla osób nagrywanych w Vicon Blade. Ponieważ, osoby te miały świadomość, że ich chód będzie potem analizowany, więc podczas nagrań można było zauważyć, że czasami poruszają się bardzo sztywno czy nienaturalnie.

W śród osób chorych, jakość klasyfikacji była zależna od głównego schorzenia. Wynika to z faktu, iż różne schorzenia, różnie objawiają się w chodzie. Osoby, które cierpiały na dodatkowe schorzenia jak osteopatia, haluksy, cukrzyca, zwyrodnienie stawów kolanowych, zwyrodnienie kręgosłupa w odcinku szyjnym, czy bóle kończyn górnych, były prawidłowo klasyfikowane.

Dla obu z wymienionych akcji sieci CNN popełniały znacznie więcej pomyłek. Wartość PBO dla większości osób była wyższa dla sieci CNN niż dla sieci LSTM.

5.5.2 Akcje statyczne

Za akcje statyczne można uznać akcje - stanie, obroty i schylanie się. W przypadku tych akcji obie sieci poradziły sobie bardzo dobrze. Rozmiar wektora wejściowego nie miał większego wpływu, na jakość klasyfikacji. Dla tych klas sieci CNN uzyskały nieznaczne lepsze wyniki niż sieci LSTM.

5.5.3 Akcje niebezpieczne

Sieci CNN znacznie lepiej poradziły sobie z prawidłową klasyfikacją akcji wykonanych przez zawodników Taekwondo - uderzenia, oraz kopnięć wysokiego i niskiego. Błędy w tych akcjach występowały, gdy wykonywali je zawodnicy Karate. Również najwięcej pomyłek dochodziło przy akcji wykonywanej tylko przez zawodników Karate - kopnięcie niskie.

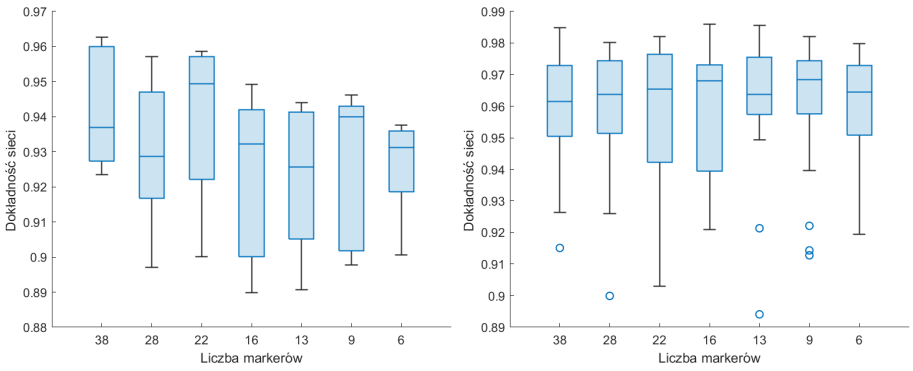
Sieci LSTM również częściej błędnie klasyfikowały zawodników Karate, jednakże błędy zdarzały się też dla zawodników Taekwondo. Dodatkowo zdarzały się osoby, których wszystkie wykonania danej techniki były błędnie klasyfikowane, co nie miało miejsca przy sieciach CNN.

Powody błędnych klasyfikacji dla obu sieci są podobne. Każda z technik wykonywana była w kilku wariantach - w powietrze lub cel (tarczę i/lub deskę). Błędy dotyczyły przede wszystkim jednego z tych wariantów. W przypadku kopnięcia niskiego brak celu powodował, iż kopnięcie było wyższe. Dla pozostałych kopnięć ustawienie celu, a zwłaszcza jego wysokość wpływała najmocniej ja klasyfikację. Co warte zaznaczenia, wbrew początkowym obawą sieci dość dobrze rozróżniały kopnięcie wysokie boczne od kopnięcia wysokiego prostego.

5.5.4 Wnioski końcowe dla sieci 3D

Ponieważ największa liczba błędów dotyczyła chodu osób zdrowych i chodu osób chorych, postanowiono ujednolicić te grupy. Na rysunku 58 przedstawiono wykresy pudełkowe dokładności sieci LSTM i CNN dla różnych rozmiarów wektora wejściowego, w których wszystkie chody zostały potraktowane, jako jedna akcja. Przy takim podziale sieci CNN zaczynają przewyższać sieci LSTM.

Pomimo tak znaczącej poprawy, na potrzeby eksperymentów z danymi dwuwymiarowymi postanowiono zachować podział chodów na dwie osobne grupy.



(a) (b)

Rysunek 58: Wykres pudełkowy dla dokładności (a) dwukierunkowej sieci LSTM, (b) sieci CNN, po zgrupowaniu akcji Chód

6 Wpływ rzutowania 2D na dokładność klasyfikacji zachowań postaci ludzkiej

6.1 Sformułowanie zadania

Głównym celem eksperymentów omawianych w niniejszym rozdziale jest określenie granicy rozpoznawalności poszczególnych akcji prostych. Do ich klasyfikacji ponownie wykorzystano dwa rodzaje głębokich sieci neuronowych - dwukierunkową sieć LSTM i jednowymiarową sieć CNN. Dodatkowym celem jest, podobnie jak w przypadku danych 3D, określenie optymalnego rozmiaru wektora wejściowego. Sposób redukcji markerów na wejściu pozostał taki sam jak w przypadku danych trójwymiarowych.

Do realizacji powyższy celów, dla każdego z 46 położeń wirtualnej kamery utworzono oba rodzaje sieci, które ponownie testowane były za pomocą 5-krotnej walidacji krzyżowej dla każdego rozmiaru wektora wejściowego. Ze względu na ograniczenia czasowe, tym razem walidacja nie była dodatkowo powtarzana.

Ocenie ponownie podlegała ogólna, jakość klasyfikacji, częstotliwość poprawnych i błędnych klasyfikacji w ramach poszczególnych klas, oraz miary PBA i PBO. Wartości te zostały porównane w ramach danej akcji dla wszystkich ustawień kamery. Została również przeprowadzona szczegółowa analiza mająca na celu sprawdzenie, czy w ramach danej akcji mylone są dane tych samych osób, co w przestrzeni trójwymiarowej. Tak samo czy na jakość klasyfikacji miały wpływ takie czynniki jak wykorzystane oprogramowanie, płeć, wiek, wzrost czy nasilenie objawów chorobowych.

6.2 Struktury wybranych sieci wraz z opisem hiper parametrów

Podobnie jak w przypadku eksperymentów na danych trójwymiarowych, w celu optymalnego doboru hiperparametrów sieci przeprowadzono optymalizacje bayerowską. Jednakże ze względu na sporą liczbę ustawień wirtualnej kamery optymalizacja ta została przeprowadzona w uproszczony sposób. Losowo wybrano kilka punktów na kopule, które nie znajdowały się koło siebie. Następnie przeprowadzano optymalizację na ograniczonym podzbiorze danych.

Optymalizacji podlegały te same hiperparametry, co w poprzednich eksperymentach:

- głębokość sieci (od 1 do 5 ukrytych warstw)
- prędkość uczenia się (0.0001 - 0.1)
- przycięcie gradientu (1-inf)

Dla sieci LSTM optymalizowano również liczbę komórek na poszczególnych warstwach (od 10 do 200). Dla sieci CNN liczbę (10-100) oraz rozmiar (2-20) filtrów. Ponownie przyjęto założenie, że przy sieciach CNN liczba filtrów ulega podwojeniu na kolejnych warstwach.

6.2.1 Sieci LSTM

Tak samo jak we wcześniejszych eksperymentach, rozmiar wektora wejściowego najbardziej wpływał, na jakość klasyfikacji, a tym samym wartości hiper-

parametrów. Wartości hiperparametrów dla części punktów ponownie oscylowały wokół podobnych wartości. Dla pozostałych wartości te były skrajnie różne, aczkolwiek w przypadku tych punktów, jakość sieci była bardzo słaba. Ma to związek z tym, iż dla pewnych ustawień kamery sieć klasyfikowała wszystkie rodzaje kopnięć, jako jedno.

Znalezienie optymalnych hiperparametrów dla każdego punktu, dla każdej wielkości wektora wejściowego jest możliwe, ale jednocześnie wymagałoby bardzo dużo czasu. Zostało wybranych 46 punktów, dla każdego punktu należałoby przeprowadzić optymalizację pod 7 różnych rozmiarów wektora wejściowego, co daje 322 sieci do optymalizacji. Czas potrzebny na tą optymalizację liczony byłby w miesiącach. Dlatego też, podobnie jak w przypadku sieci operujących na danych trójwymiarowych, postanowiono porównać wartości hiperparametrów uzyskanych w poszczególnych próbach a następnie je uśrednić. Rezultaty okazały się być bardzo zbliżone do poprzednich:

- głębokość sieci – 2 warstwy ukryte
- liczba ukrytych jednostek na warstwie – 100/50
- prędkość uczenia się – 0.001
- przycięcie gradientu – 1

Schemat zaproponowanej sieci przedstawiono na rysunku 59.

6.2.2 Sieci CNN

W przypadku sieci CNN wpływ rozmiaru wektora wejściowego był nieco mniejszy. Wartości hiperparametrów dla losowych punktów były bardziej zbliżone do siebie niż w przypadku sieci LSTM. Dodatkowo oscylowały one wokół podobnych liczb jak dla sieci utworzonych dla danych trójwymiarowych. Dlatego też, postanowiono, że wartości hiperparametrów pozostaną takie same jak przy wyżej wspomnianych sieciach:

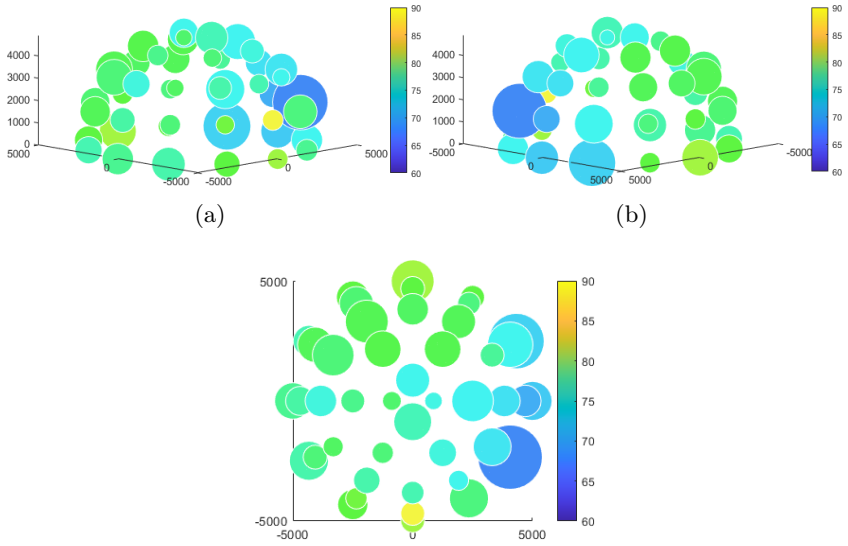
- głębokość sieci – 2 warstwy ukryte
- liczba filtrów na warstwie – 32/64
- rozmiar filtra - 5
- prędkość uczenia się – 0.001
- przycięcie gradientu – inf

6.3 Metoda wizualizacji dokładności klasyfikacji

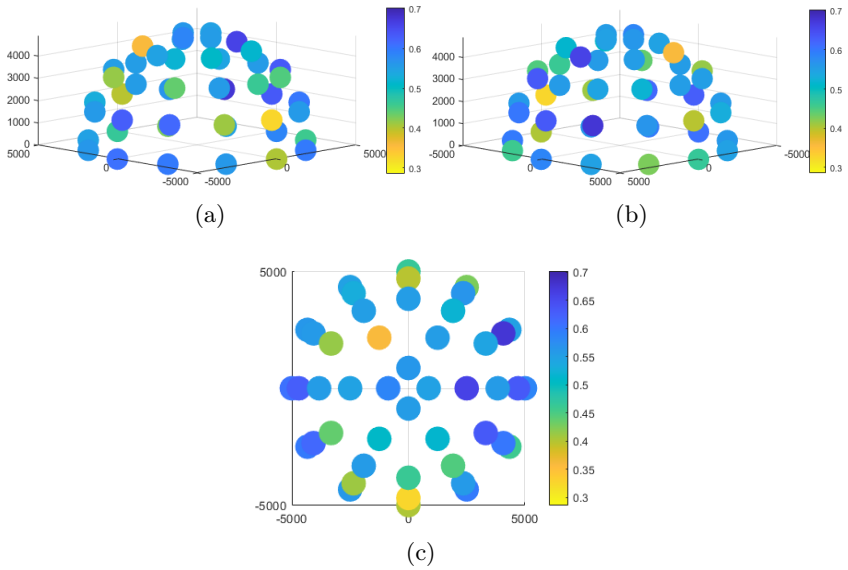
W poprzednim rozdziale dokładności sieci, oraz wartości PBO i PBA przedstawione zostały za pomocą wykresów pudełkowych i histogramów. W przestrzeni dwuwymiarowej, ten rodzaj wizualizacji, mimo iż możliwy byłby nieefektywny - liczba utworzonych wykresów była by zbyt duża. Dlatego też, w celu prostego i szybkiego wizualnego podsumowania wyników opracowano prezentacje w formie specjalych kopuł, gdzie każdy punkt na kopule odpowiada jednemu z położeń wirtualnej kamery. Kolory i rozmiar tych punktów różnią się w zależności od wartości, jaką wizualizują.

W wizualizacji wpływu rozmiaru wektora wejściowego, na jakość klasyfikacji kolor każdego punktu odpowiada średniej, dokładności klasyfikacji. Punkty z niską średnią dokładnością (około 50%) mają kolor ciemnoniebieski, a sieci z wysoką (90%) jasnożółty. Rozmiar punktu natomiast podyktowany jest różnicą pomiędzy najlepszym a najgorszym wynikiem w danym punkcie. Im większa rozrzut w wynikach tym większy rozmiar punktu. Na rysunku 60 przedstawiono kolejno przykładowe kopuły dla jakości klasyfikacji widziane z trzech różnych perspektyw. Pierwsze dwie (a i b) to widok z przeciwległych boków, trzeci (c) to widok z góry.

W przypadku wartości PBA rozmiar punktów jest stały, zmianie ulega tylko kolor. W celu łatwiejszej wizualnej oceny postanowiono zachować spójność kolorystyczną. Sieci z wyższą wartością PBA, czyli takie, w których częściej dochodziło do pomyłek, zostały zaznaczone na ciemnoniebiesko. Im mniejsza wartość PBA tym kolor punktu jaśniejszy. Przy czym należy mieć na uwadze, iż wartości PBA różnią się pomiędzy poszczególnymi akcjami - dla akcji chód pomyłki dotyczyły prawie 70% przejść, podczas gdy dla stania nie było to więcej jak 10%. Dlatego też skala kolorów jest dobierana dynamicznie dla każdej akcji. Na rysunku 61 przedstawiono kolejno przykładowe kopuły dla wartości PBA dla akcji chód osób zdrowych widziane z trzech różnych perspektyw. Pierwsze dwie (a i b) to widok z przeciwległych boków, trzeci (c) to widok z góry.



Rysunek 60: Kopuła dokładności sieci CNN dla wszystkich położeń wirtualnej kamery dla 38 znaczników w wektorze wejściowym, widziana z różnych perspektywy (a) z boku, (b) z boku, (c) od góry.

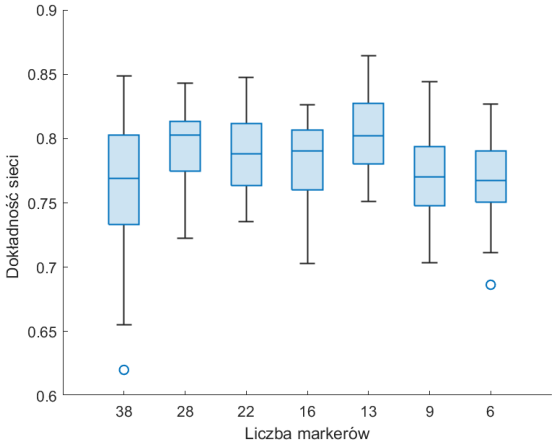


Rysunek 61: Kopuła wartości PBA sieci CNN dla akcji chód osób zdrowych, dla 38 znaczników wektorze w wejściowym, widziana z różnych perspektyw (a) z boku, (b) z boku, (c) od góry.

6.4 Klasyfikacja z wykorzystaniem sieci LSTM

6.4.1 Wpływ rozmiaru wektora wejściowego na dokładność sieci

Podobnie jak w przypadku sieci utworzonych dla danych trójwymiarowych, rozmiar wektora wejściowego w znaczący sposób wpływał zarówno na dokładność sieci jak i jej spójność. Na rysunku 62 przedstawiono pudełkowy wykres dla średniej dokładności wszystkich sieci LSTM, dla różnej liczby markerów na wejściu, dla wszystkich położeń wirtualnej kamery.



Rysunek 62: Wykres pudełkowy dla średniej dokładności wszystkich sieci LSTM utworzonych dla każdego położenia wirtualnej kamery dla różnej liczby markerów w wektorze wejściowym

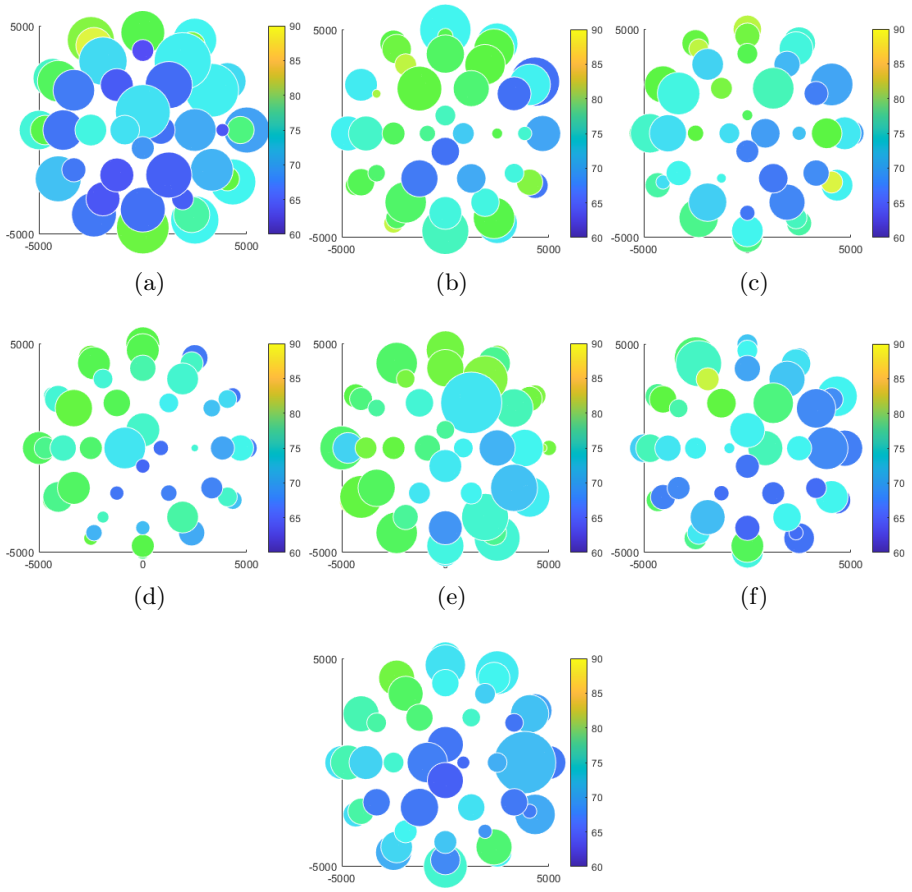
Można zauważyć podobną charakterystykę jak przy sieciach utworzonych dla danych trójwymiarowych. Po pierwszej największej redukcji rozmiaru wektora wejściowego następuje lekka poprawa, w jakości sieci. Natomiast każda kolejna redukcja sukcesywnie obniża ogólną dokładność sieci. Wyjątek stanowiły sieci utworzone dla wektora wejściowego składającego się z 13 znaczników. Uzyskały one średnio lepsze wyniki. Spójność wyników zwracanych przez sieci jest najgorsza dla wektora wejściowego składającego się z 38 znaczników - różnica pomiędzy najlepszą a najgorszą siecią wynosi 30%. Dla pozostałych rozmiarów wektora wejściowego wynosiła ona zwykle około 10%.

Na rysunkach 63,64 i 65 przedstawiono kopuły dokładności sieci dla różnych rozmiarów wektora wejściowego. Dodatkowo na rysunku 66 przedstawiono wykresy pudełkowe dokładności sieci ponownie z uwzględnieniem różnych położeń kamery oraz liczby markerów na wejściu (dostępne w pełnej rozdzielczości w dodatku B). Rysunki te doskonale wizualizują kilka zależności.

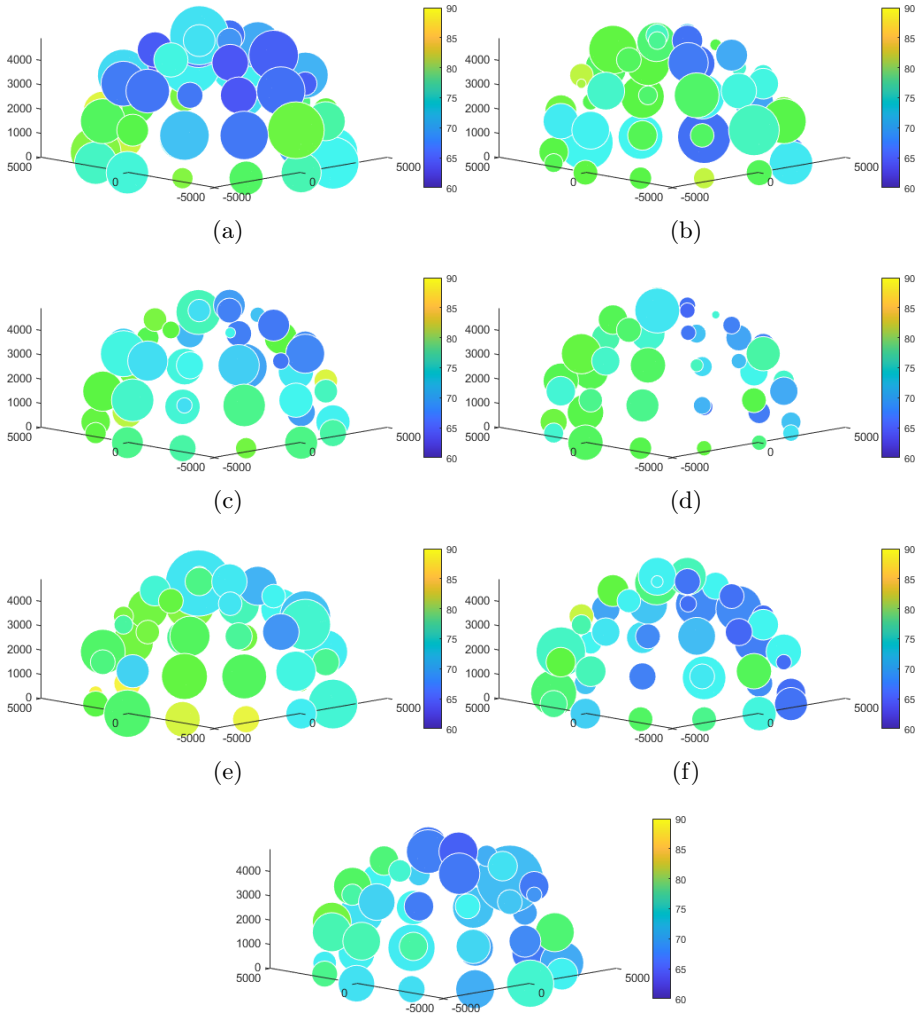
Po pierwsze, zgodnie z początkowymi przypuszczeniami, niezależnie od wielkości wektora wejściowego istnieją punkty, które uzyskiwały za każdym razem dużo lepsze/gorsze rezultaty. Akcje widziane z przodu lub tyłu charakteryzowały się największą ogólną dokładnością. Dla ustawień bocznych, dokładność ta była znacznie mniejsza, zwłaszcza dla prawej strony kopuły. Zależność ta ma ogromny wpływ na omawiany wcześniej wykres średniej dokładności (rys. 62). W przypadku niektórych sieci możemy zauważyć, że średnią w znaczący sposób

zaniża kilka punktów.

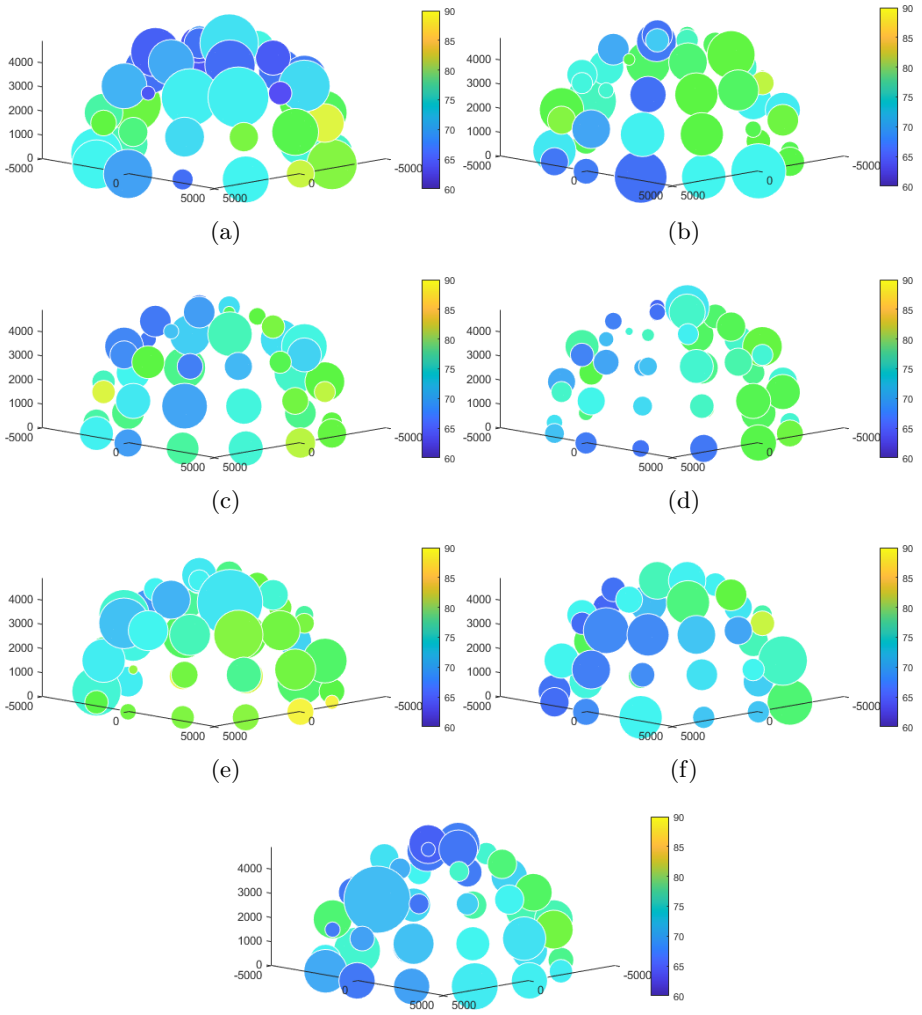
Kolejna zależność dotyczy spójności wyników sieci. Niższa średnia dokładność powiązana jest najczęściej z większą różnicą pomiędzy najlepszą a najgorszą siecią w danym punkcie. Odwrotna sytuacja też zachodzi dość często - sieci z wyższą dokładnością są zazwyczaj bardziej spójne. Dodatkowo wahania w dokładności dla poszczególnych położeń wirtualnej kamery są zbliżone dla różnych rozmiarów wektora wejściowego. Zdarzają się też wyjątki, gdy tylko dla jednego rozmiaru wektora wejściowego sieci uzyskuję bardzo dobre, zbliżone do siebie rezultaty.



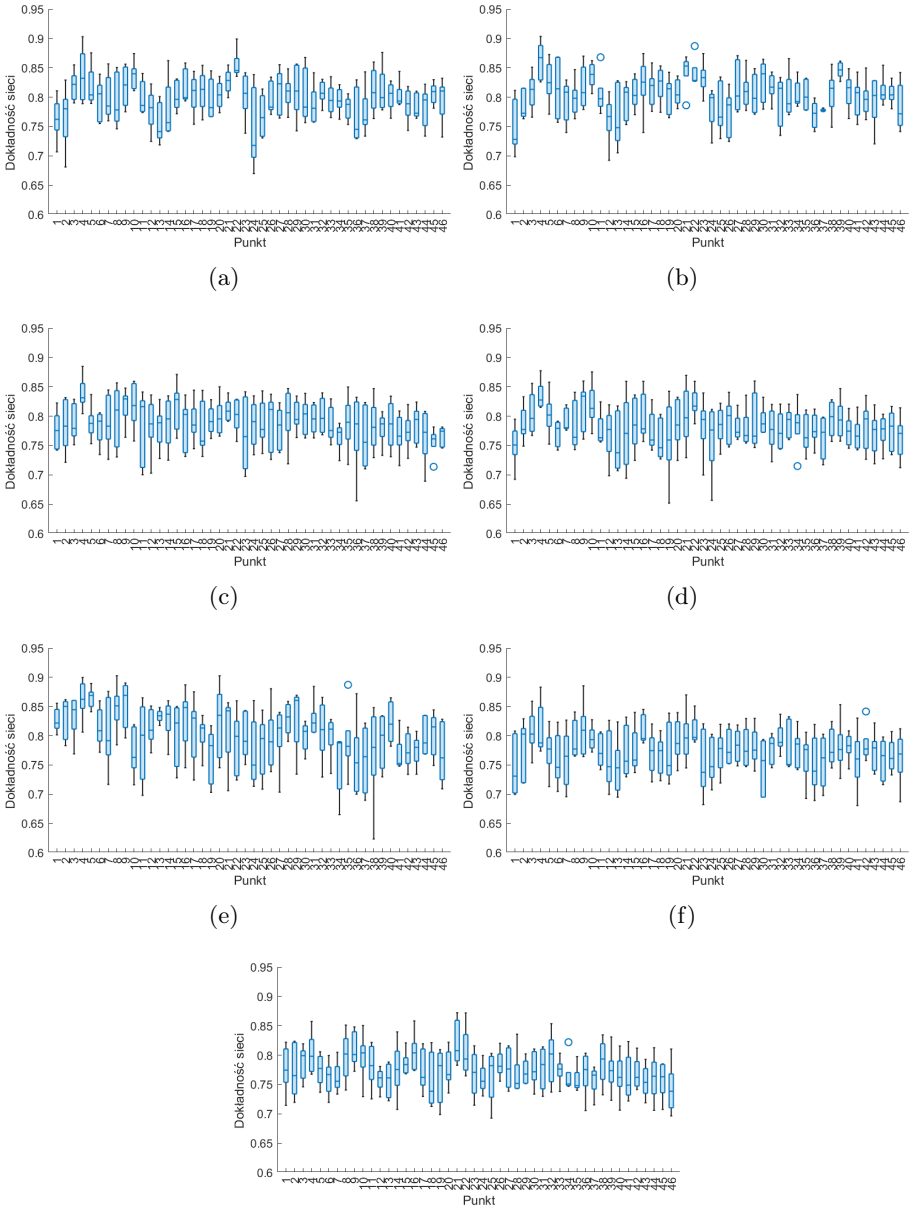
Rysunek 63: Dokładność klasyfikacji sieci LSTM dla wszystkich położeń wirtualnej kamery dla różnej liczby znaczników w wektorze wejściowym: (a) 38 znaczników, (b) 28 znaczników, (c) 22 znaczniki, (d) 16 znaczników, (e) 13 znaczników, (f) 9 znaczników, (g) 6 znaczników, widziana z góry



Rysunek 64: Dokładność klasyfikacji sieci LSTM dla wszystkich położeń wirtualnej kamery dla różnej liczby znaczników w wektorze wejściowym: (a) 38 znaczników, (b) 28 znaczników, (c) 22 znaczniki, (d) 16 znaczników, (e) 13 znaczników, (f) 9 znaczników, (g) 6 znacznikó, widziana z perspektywy połodniowo-zachodniej.



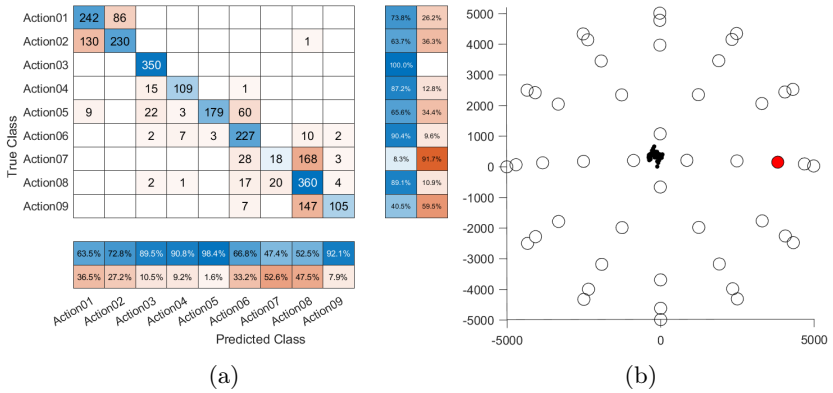
Rysunek 65: Dokładność klasyfikacji sieci LSTM dla wszystkich położeń wirtualnej kamery dla różnej liczby znaczników w wektorze wejściowym: (a) 38 znaczników, (b) 28 znaczników, (c) 22 znaczniki, (d) 16 znaczników, (e) 13 znaczników, (f) 9 znaczników, (g) 6 znaczników, widziana z perspektywy północno-wschodniej.



Rysunek 66: Wykres pudełkowy dokładności sieci dla wszystkich położeń wirtualnej kamery, dla różnej liczby znaczników w wektorze wejściowym: (a) 38 znaczników, (b) 28 znaczników, (c) 22 znaczniki, (d) 16 znaczników, (e) 13 znaczników, (f) 9 znaczników, (g) 6 znaczników.

Analiza macierzy pomyłek wykazała, iż ponownie do błędów dochodziło w obrębie podobnych akcji. Na rysunku 68 przedstawiono macierze pomyłek dla jednego z punktów, w którym dokładność klasyfikacji wahała się od 80% do 90%, dla 13 markerów w wektorze wejściowym. Możemy zauważyć, iż macierze te wyglądają podobnie do tych uzyskanych dla danych trójwymiarowych. Do pomyłek najczęściej dochodzi w ramach obu akcji chód, oraz poszczególnych kopnięć.

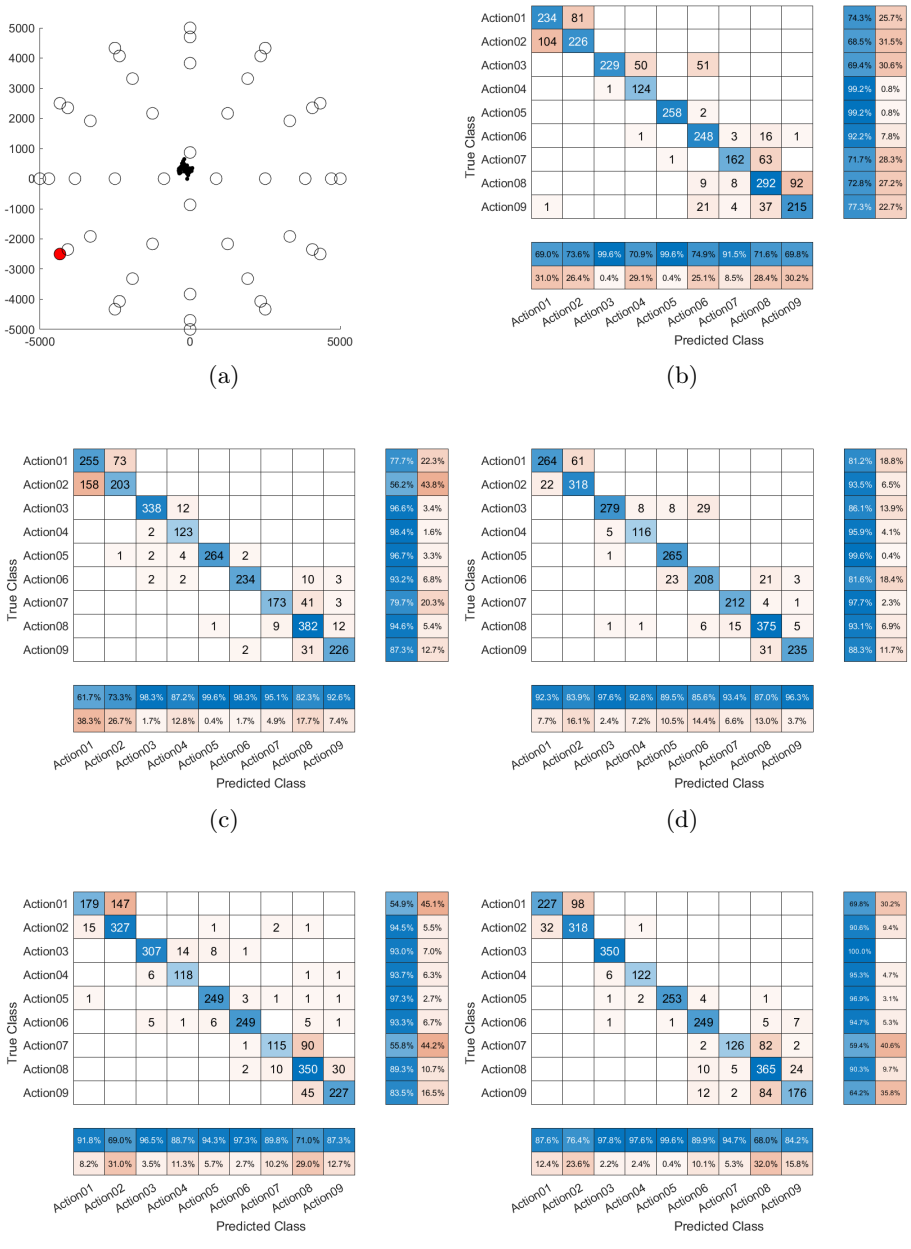
Na rysunku 67 przedstawiono macierz pomyłek dla punktu, który uzyskał najgorsze rezultaty - dokładność na poziomie 70% lub mniej. Można zauważyć, iż w tym przypadku dokładność sieci w znaczący sposób zaniżyła akcja kopnięcie niskie, oraz kopnięcie wysokie boczne. Niemal wszystkie wykonania obu wymienionych rodzajów kopnięć zostały sklasyfikowane, jako kopnięcie wysokie proste.



(a) (b)

Rysunek 67: Przykładowa macierz pomyłek (a) dla wybranego położenia wirtualnej kamery (b) dla 13 markerów w wektorze wejściowym

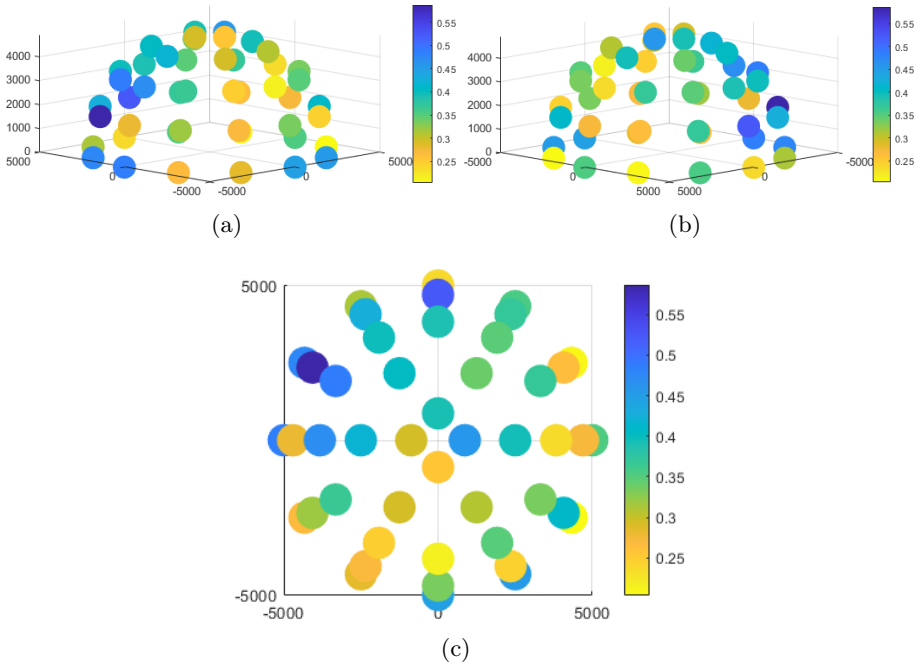
Poszczególne rodzaje kopnięć były mylone pomiędzy sobą tylko w przypadku niektórych położeń wirtualnej kamery. Akcje chód były natomiast mylone między sobą zawsze w większym lub mniejszym stopniu, niezależnie od położenia wirtualnej kamery, oraz liczby markerów na wejściu. Dlatego też postanowiono ponownie wyeliminować wpływ błędów akcji chód na ogólną, dokładność klasyfikacji. Po tej eliminacji okazało się, iż najlepsze rezultaty osiągnięto dla sieci utworzonych dla 13 markerów w wektorze wejściowym. Zarówno mniejsza jak i dużo większa liczba znaczników na wejściu powodowała większą liczbę pomyłek pomiędzy poszczególnymi klasami. Dlatego też, dalsza bardziej szczegółowa analiza wpływu położenia wirtualnej kamery, na czułość klasyfikacji zostanie przeprowadzona z pominięciem wpływu wielkości wektora wejściowego.



Rysunek 68: Przykładowe macierze pomyłek dla kolejnych walidacji (b-f) dla wybranego położenia wirtualnej kamery (a), dla 13 markerów w wektorze wejściowym

6.4.2 Akcja Chód osób zdrowych

Akcja chód osób zdrowych w zależności od położenia wirtualnej kamery cechowała się czułością na poziomie od 41% do 80%. Na rysunku 69 przedstawiono kopułę wartości PBA dla tej akcji. Do błędów częściej dochodziło, gdy kamera ustawiona była za lub z prawej strony osoby. Dodatkowo im wyżej położona była kamera tym częściej dochodziło do pomyłek.

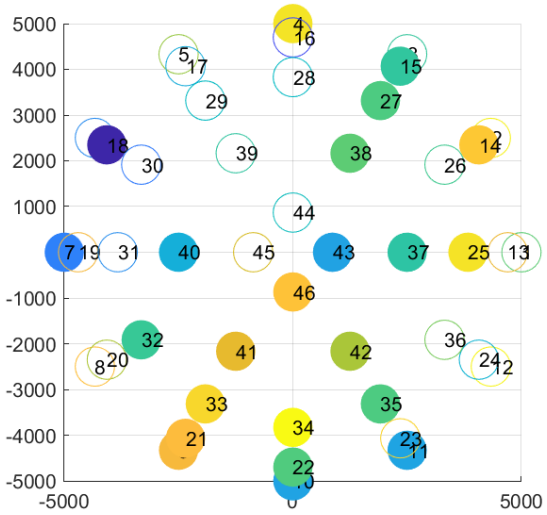


Rysunek 69: Kopuła wartości PBA sieci LSTM, dla akcji chód osób zdrowych, widziana z trzech różnych perspektyw (a) z boku, (b) z boku, (c) od góry.

Podobnie jak w przypadku danych trójwymiarowych, akcja ta mylona była głównie z akcją chód osób chorych. Dla kilku położeń wirtualnej kamery dochodziło do pomyłek z innymi klasami. Na rysunku 70 ponownie przedstawiono kopułę wartości PBA widzianą z góry. Na rysunku tym punkty, dla których dochodziło do pomyłek tylko z akcją chód osób chorych pozostały puste, a te gdzie dochodziło również do pomyłek z innymi klasami zostały wypełnione. Dodatkowo w tabeli 6 przedstawiono listę punktów, w których dochodziło do pomyłek z daną akcją, wraz z łączną liczbą pomyłek we wszystkich wymienionych punktach. Do pomyłek z innymi akcjami dochodziło sporadycznie - średnio 1-2 przejścia były klasyfikowane, jako jedna z pozostałych akcji. W punkcie 27 dochodziło do pomyłek z największą liczbą klas.

Podobnie jak w przypadku danych 3D, dla każdego położenia wirtualnej kamery znalazły się osoby, których wszystkie przejścia były poprawnie/błędnie klasyfikowane. W większości były to te same osoby. Zdarzały się jednak sytuację, że dana osoba z danej perspektywy klasyfikowana była zawsze prawidłowo, a z innej zawsze lub prawie zawsze niepoprawnie. Znalazły się też osoby, których

chód, niezależnie od położenia wirtualnej kamery zawsze był tak samo klasyfikowany.



Rysunek 70: Kopuła wartości PBA sieci LSTM, dla akcji chód osób zdrowych, widziana z góry, wraz z numerami punktów.

Wpływ wykorzystywanego do nagrań oprogramowania w nieco mniejszym stopniu wpływa, na dokładność klasyfikacji. Ponownie większość osób poprawnie klasyfikowanych nagrywanych było za pomocą oprogramowania Vicon Blade. Jednakże w grupie 6 osób, które zawsze, niezależnie od położenia wirtualnej kamery były poprawnie klasyfikowane znalazła się jedna osoba nagrywana za pomocą Vicon Nexus. Osoby te były również zawsze poprawnie identyfikowane w przestrzeni trójwymiarowej.

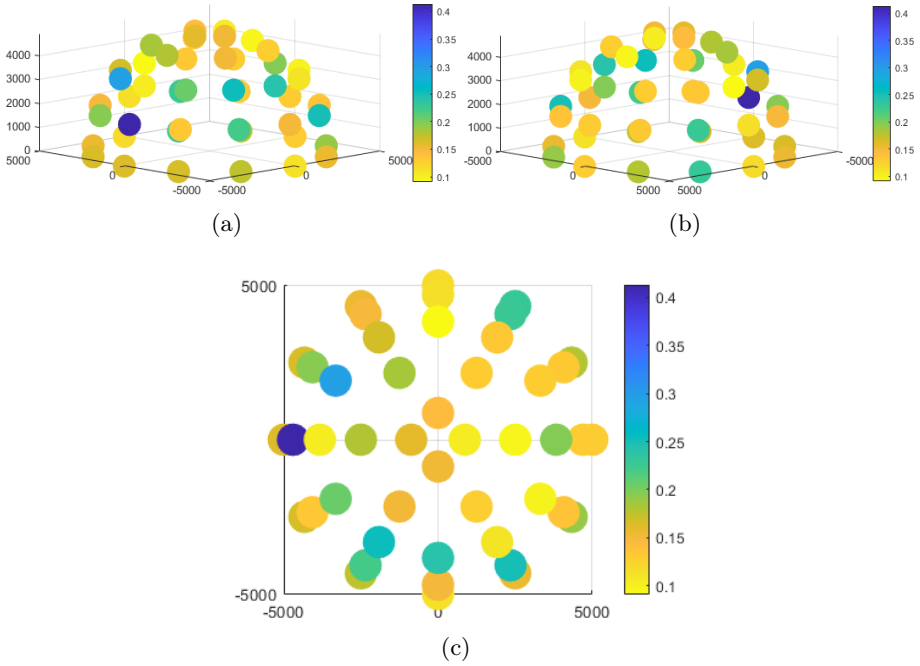
Po przeciwnej stronie - wśród osób zawsze błędnie klasyfikowanych znalazły się w większości osoby nagrywane w Nexusie. Wśród 11 osób, zawsze błędnie identyfikowanych, niezależnie od położenia wirtualnej kamery były tylko osoby nagrywane w tym oprogramowaniu. Co warte zaznaczenia, tylko jedna z tych osób była zawsze błędnie klasyfikowana w przestrzeni trójwymiarowej, przejścia pozostałych osób były błędnie identyfikowane w 70-90%.

Tabela 6: Lista punktów, dla których dochodziło do pomyłek akcji chód osób zdrowych (na 74474 klasyfikacji) z innymi akcjami

Akcja	Punkty	Pomyłki
Stanie	15 32	7
Obroty	7 21	6
Schylanie się	4 27 32 40	12
Uderzenie	27	2
Kopnięcie niskie	15 27	2
Kopnięcie w. proste	9 10 11 14 21 22 34 38 41 42	23
Kopnięcie w. boczne	11 27 35 38 43	36

6.4.3 Akcja Chód osób chorych

Akcja chód osób chorych, podobnie jak w przypadku danych trójwymiarowych cechowała się większą rozpoznawalnością niż chód osób zdrowych. W zależności od położenia wirtualnej kamery czułość tej akcji wahała się od 58% do 90%. Na rysunku 71 przedstawiono kopułę wartości PBA dla omawianej akcji. Możemy zauważyć korelację pomiędzy wartościami PBA akcji chód osób zdrowych i chód osób chorych. Punkty, dla których wartość ta była wysoka dla jednej z wymienionych akcji, dla drugiej była niska i odwrotnie. Do pomyłek dochodziło w głównej mierze z akcją chód osób zdrowych. Jednakże w każdym punkcie dochodziło, do co najmniej jednej pomyłki z innymi klasami.



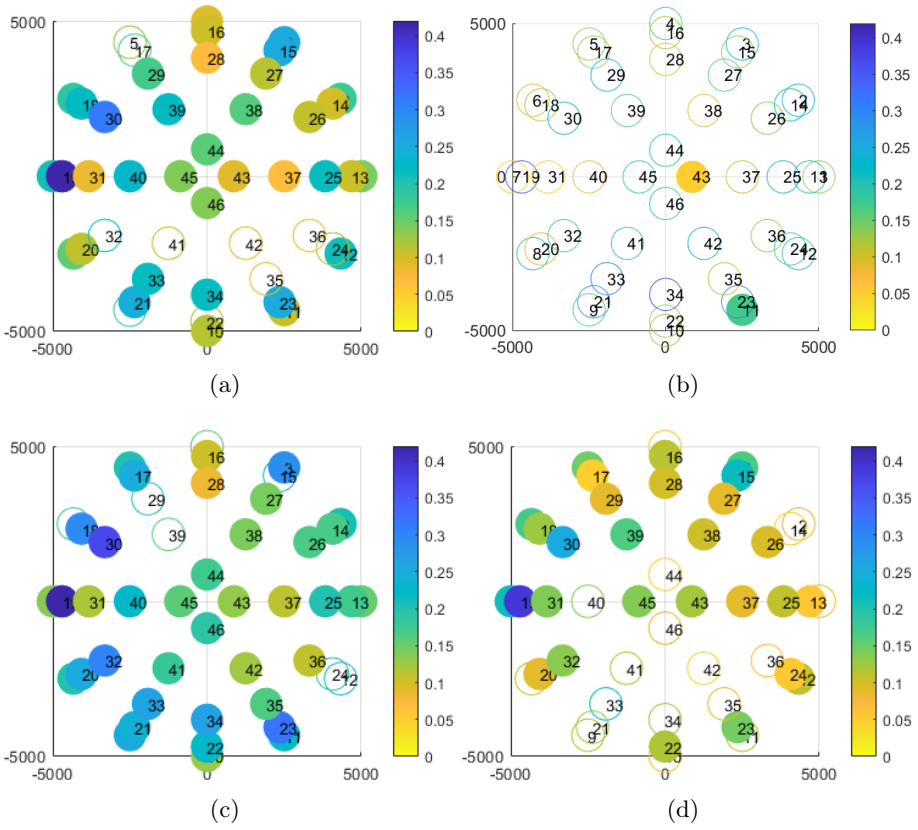
Rysunek 71: Kopuła wartości PBA sieci LSTM, dla akcji chód osób chorych, widziana z trzech różnych perspektyw (a) z boku, (b) z boku, (c) od góry.

W zależności od położenia wirtualnej kamery, osoby były w różnym stopniu poprawnie klasyfikowane. W skrajnych przypadkach dla jednego położenia wartość PBO była 1, dla innego 0. W grupie osób, dla których sieci zawsze dawały poprawne wyniki znalazły się 4 osoby - pacjenci ze zwyrodnieniem kręgosłupa. Natomiast wśród osób zawsze błędnie klasyfikowanych, jako zdrowe znalazło się 3 pacjentów: jedne po endoprotezoplastyce obu stawów biodrowych, oraz dwóch pacjentów z chorobą Parkinsona.

Podobnie jak w przestrzeni trójwymiarowej, przeprowadzono analizę wpływu danej jednostki chorobowej, na czułość klasyfikacji. Na rysunku 72 przedstawiono kopuły PBA, widziane z góry, dla poszczególnych jednostek chorobowych. Punkty puste, to punkty, w których dochodziło do pomyłek tylko z akcją chód osób zdrowych, w punktach wypełnionych kolorem dochodziło również do po-

myłek z innymi akcjami.

Pacjenci po endoprotezoplastyce stawu biodrowego byli myleni głównie z chodem osób zdrowych. Do pomyłek z innymi klasami dochodziło zwykle w przypadku jednego czy dwóch przejść, co jest widoczne w tabeli 7, gdzie przedstawiono listę punktów, w których dochodziło do pomyłek z daną akcją, wraz z informacją o łącznej liczbie pomyłek w tych punktach.



Rysunek 72: Kopuła wartość PBA sieci LSTM, dla akcji chód osób chorych, widziana z góry, wraz z numerami punktów, dla różnych jednostek chorobowych (a) endoprotezoplastyka stawu biodrowego, (b) zwyrodnienie kręgosłupa, (c) choroba Parkinsona, (d) udar niedokrwienny mózgu.

Dla pewnych ułożeń wirtualnej kamery można zaobserwować podobne korelacje, pomiędzy jakością klasyfikacji a stanem ogólnym pacjenta, jak w przypadku danych trójwymiarowych. Pacjenci po endoprotezoplastyce obu stawów biodrowych byli częściej klasyfikowani, jako zdrowi, gdy kamera znajdowała się z ich prawej lub lewej strony. Najlepsze rezultaty uzyskano, gdy kamera znajdowała przed lub za pacjentem. Wynika to prawdopodobnie z faktu, iż osoby po takim zabiegu mają najczęściej ograniczony ruch w stawie biodrowym w płaszczyźnie czołowej. Podczas rzutowania z innych perspektyw, ograniczenie to może nie być tak dobrze widoczne.

Osoby ze zwyrodnieniem kręgosłupa, poza jedną lokalizacją wirtualnej kamery, myleni byli tylko z chodem osób zdrowych. Wyjątek stanowiły punkty 11 i 43, w których doszło łącznie do 4 pomyłek z akcją kopnięcie wysokie boczne. W tej grupie pacjentów do pomyłek dochodziło częściej, gdy wirtualna kamera znajdowała się na wprost pacjenta. Najlepsze rezultaty uzyskano dla projekcji bocznych, gdzie wartości PBA wahały się od 0,05 do 0,12.

Chód osób z chorobą Parkinsona mylony był nie tylko z chodem osób zdrowych, ale również ze wszystkimi pozostałymi klasami, dla większości położeń wirtualnej kamery. W tabeli 8 przedstawiono zestawienie punktów, w których dochodziło do pomyłek z daną klasą. Ponownie, pomyłki te dotyczyły głównie 1/2 przejść w danym punkcie. Co warte zaznaczenia, do pomyłek z akcją obroty dochodziło w prawie połowie położeń wirtualnej kamery. W odróżnieniu od pozostałych schorzeń, rozpoznawalność tej akcji była bardzo zbliżona dla wszystkich położeń wirtualnej kamery. Możemy też zauważyć znacznie mniejsze różnice w wartości PBO dla danej osoby uzyskanej w różnych punktach.

Ostatnim z omawianych schorzeń jest stan po przebytym udarze niedokrwiennym mózgu. Akcja ta niezależnie od punktu charakteryzowała się dużą rozpoznawalnością. Dla większości położeń wirtualnej kamery wartość PBA wahała się między 0,05 a 0,12. Wyjątek stanowił jeden punkt - 19, gdzie wartość ta wynosiła 0,4. Wysoka rozpoznawalność tej akcji może mieć związek z tym, iż schorzenie to wpływa na całą motorykę pacjenta. Ponownie w niektórych punktach dochodziło do pomyłek z różnymi akcjami. Zestawienie tych pomyłek zostało przedstawione w tabeli 9.

Akcja chód osób chorych cechowała się dość dobrą rozpoznawalnością. Można zauważyć korelację pomiędzy jednostkami chorobowymi i położeniem wirtualnej kamery, na czułością klasyfikacji. Wpływ niektórych schorzeń na chód jest bardziej widoczny z pewnych perspektyw, zwłaszcza, jeśli upośledzają one ruch kończyn tylko w jednej płaszczyźnie.

Dla każdej jednostki chorobowej ponownie znalazły się osoby, które w jednym punkcie były zawsze poprawnie klasyfikowane a w innym zawsze błędnie. W większości przypadków, osoby, które w przestrzeni trójwymiarowej były błędnie klasyfikowane, również dla większości położeń wirtualnej kamery osiągały dużo gorsze wyniki. Podobnie osoby zawsze poprawnie klasyfikowane w przestrzeni 3D dla większości punktów uzyskiwały wysoki procent rozpoznawalności.

Tabela 7: Lista punktów, dla których dochodziło do pomyłek akcji chód osób chorych (na 20654 klasyfikacji) z innymi akcjami, wśród osób po endoprotezoplastyce stawu biodrowego

Akcja	Punkty	Pomyłki
Stanie	3 15 28	6
Obroty	2 7 10	18
Schylanie się	3 8 16 20 23	9
Uderzenie	1 14 15 21 25 37 38 44 45	13
Kopnięcie niskie	7 8 11 14 16 18 19 30 31 40	23
Kopnięcie w. proste	2 3 6 7 8 12 25 26 27 28 33 38	32
Kopnięcie w. boczne	4 11 12 13 15 26 27 29 34 37 38 39 43 46	21

Tabela 8: Lista punktów, dla których dochodziło do pomyłek akcji chód osób chorych (na 12696 klasyfikacji) z innymi akcjami, wśród osób z chorobą Parkinsona

Akcja	Punkty	Pomyłki
Stanie	11 13 23 34	4
Obroty	1 2 3 7 8 9 16 17 20 21 22 25 31 32 33 40 41 42 43 45 46	30
Schylanie się	36	2
Uderzenie	2 14 37 38 41	9
Kopnięcie niskie	23 27 28 38 41	5
Kopnięcie w. proste	10 19 20 21 22 25 26 27 35 37 38 43 45	14
Kopnięcie w. boczne	1 5 18 20 26 30 31 32 40 44	13

Tabela 9: Lista punktów, dla których dochodziło do pomyłek akcji chód osób chorych (na 13018 klasyfikacji) z innymi akcjami, wśród osób po udarze

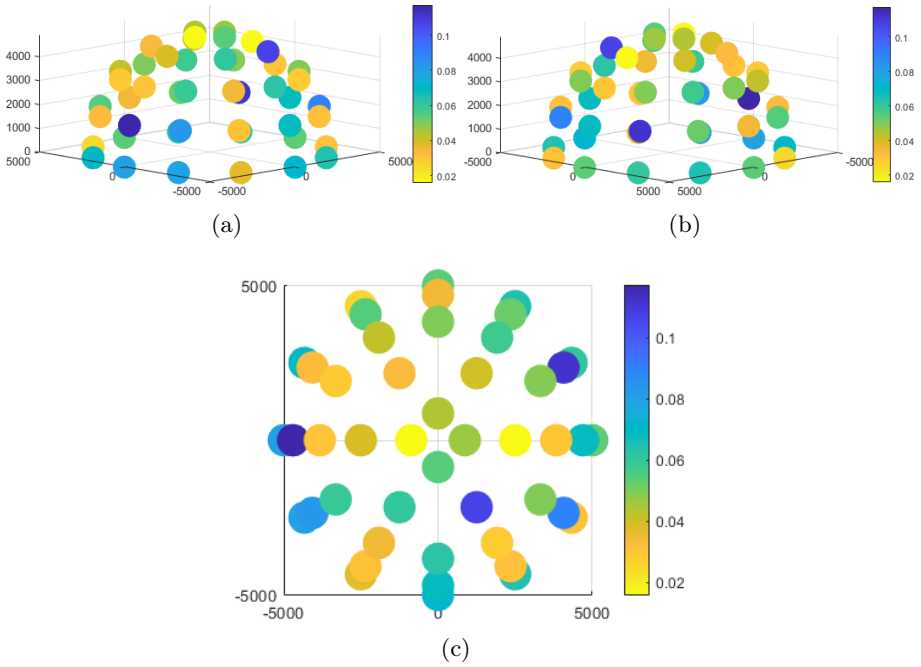
Akcja	Punkty	Pomyłki
Stanie	15 19	4
Obroty	6 7 20 32 37	15
Schylanie się	3 5 28 31 38 45	28
Uderzenie	12 16 23 25 32 37 38 43 45	12
Kopnięcie niskie	7 13 27 29 30 31 45	7
Kopnięcie w. proste	16 17 18 20 22 24 27 28 29 37 39	24
Kopnięcie w. boczne	12 13 18 26 31	9

6.4.4 Akcja Stanie

Akcja stanie, podobnie jak w przypadku danych trójwymiarowych, charakteryzowała się bardzo wysoką rozpoznawalnością. W zależności od położenia wirtualnej kamery, czułość klasyfikacji wynosiła od 88% do 98%. Na rysunku 73 przedstawiono kopułę wartości PBA dla omawianej akcji. Możemy zauważyć, iż dla punktów znajdujących się bliżej podłoża, wartości PBA są nieznacznie niższe.

Akcja ta najczęściej mylona była z pozostałymi akcjami statycznymi: obroty i schylanie się, oraz z akcją uderzenie. W tabeli 10 przedstawiono zestawienie punktów wraz z informacją, z jaką akcją i ile razy mylona była akcja stanie.

Stanie różnych osób w różnych punktach były mylone z akcją obroty. Znalazły się jednak osoby, które niezależnie od położenia wirtualnej kamery zawsze były błędnie klasyfikowane. Są to te same osoby, dla których dochodziło do pomyłek w przypadku danych trójwymiarowych. Przy czym tylko 3 osoby były praktycznie zawsze błędnie klasyfikowane. Pozostałe tylko w kilku punktach. Podobna zależność występuje dla akcji schylanie się oraz uderzenie - osoby, które były w przestrzeni 3D mylone, przynajmniej raz z daną akcją, są również mylone dla pewnych położeń wirtualnej kamery.



Rysunek 73: Kopuła wartości PBA sieci LSTM, dla akcji stanie, widziana z trzech różnych perspektyw (a) z boku, (b) z boku, (c) od góry.

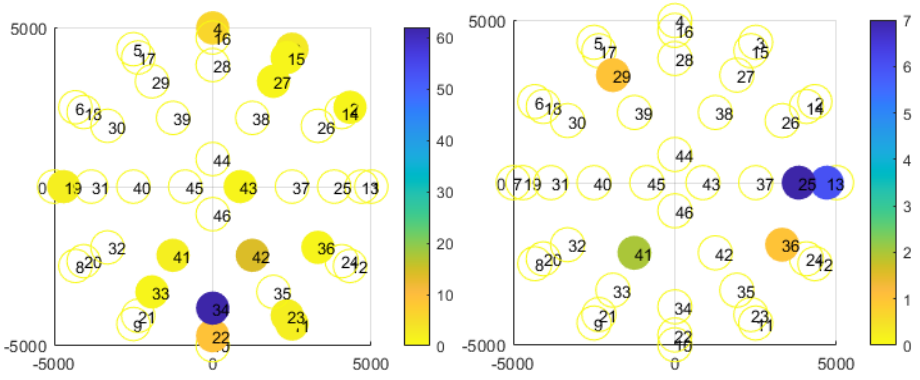
Tabela 10: Lista punktów, dla których dochodziło do pomyłek akcji stanie (na 77464 klasyfikacji) z innymi akcjami

Akcja	Punkty	Pomyłki
Chód osób zdrowych	4	1
Chód osób chorych	11 14 19 22 35 46	32
Obroty	Wszystkie	2984
Schylanie się	Wszystkie oprócz 4 17 34 35	580
Uderzenie	Wszystkie	2235
Kopnięcie niskie	1 4 25 38 44	11
Kopnięcie w. proste	2 3 4 11 15 19 22 23 27 33 34 36 41 42 43	108
Kopnięcie w. boczne	13 25 29 36 41	11

Przypuszczalna przyczyna występowania pomyłek jest podobna jak przy danych trójwymiarowych - osoby te wykonywały bardziej dynamiczne ruchy podczas stania. Ruchy te pod pewnymi kątami są lepiej widoczne, dlatego też ich, dokładność klasyfikacji zależy od ustawienia wirtualnej kamery.

Punkty, które charakteryzowały się niską wartością PBA były również punktami, w których dochodziło do pomyłek z najmniejszą liczbą innych akcji. W przypadku kamery umieszczonej po prawej stronie pacjenta, wartości PBA jest nieznacznie większa, jednakże do pomyłek w tych punktach dochodzi głównie z innymi akcjami statycznymi. W przypadku, gdy wirtualna kamera znajduje

się na wprost, lub za osobą oraz jest stosunkowo nisko, dochodzi do pomyłek z pozostałymi akcjami, głównie kopnięciem wysokim prostym. Do pomyłek tych też częściej dochodzi, gdy kamera jest z lewej strony pacjenta. Na rysunku 74 przedstawiono kopułę punktów, w której dochodziło do pomyłek z dwoma rodzajami kopnięć wysokich. Do błędów dochodziło w punktach wypełnionych, a ich kolor determinowany jest łączną liczbą pomyłek w tym punkcie.



(a) (b)

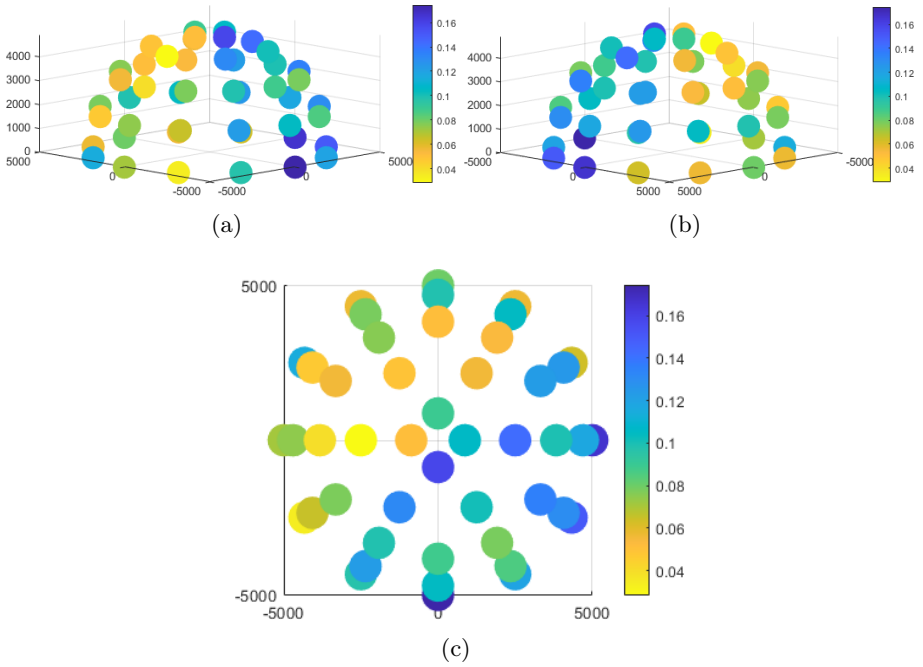
Rysunek 74: Kopuła liczby błędnych klasyfikacji wraz z ich numerami dla sieci LSTM, w których akcja stanie mylona była z akcją kopnięcie wysokie (a) proste, (b) boczne.

6.4.5 Akcja Obroty

Akcja obroty również charakteryzowała się bardzo wysoką rozpoznawalnością. Czułość klasyfikacji tej akcji w zależności od położenia wirtualnej kamery, wahała się od 82% do 97%. Niższe rezultaty uzyskiwały punkty znajdujące się z lewej strony pacjenta (rys. 75). Dodatkowo ponownie, gdy kamera znajdowała się bliżej podłoża uzyskiwano nieznacznie gorsze rezultaty.

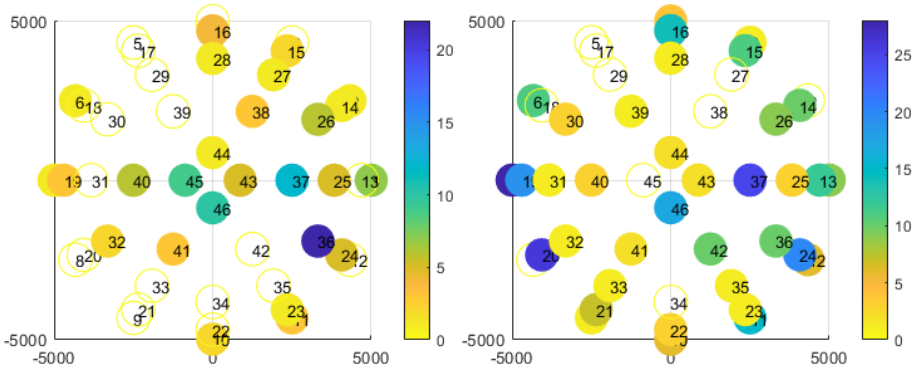
Do największej liczby pomyłek dochodziło z akcjami: stanie, uderzenie, oba kopnięcia wysokie, oraz schylanie się. W tabeli 11 przedstawiono zestawienie punktów, w których dochodziło do pomyłek z poszczególnymi klasami. Dodatkowo, na rysunkach 76 i 77 przedstawiono kopułę punktów wraz z zaznaczonymi punktami, dla poszczególnych akcji, gdzie o kolorze punktu decyduje łączna liczba pomyłek w tym punkcie.

Do pomyłek z akcjami schylanie się, uderzenie czy wszystkie rodzaje kopnięć, dochodziło znacznie częściej, gdy kamera znajdowała się nad pacjentem. Ponadto ponownie, akcje te były częściej mylone, gdy wirtualna kamera znajdowała się z lewej strony pacjenta. Co warte zaznaczenia, osoba, której obroty mylone były w przestrzeni trójwymiarowej z praktycznie wszystkimi innymi rodzajami akcji, również w przestrzeni dwuwymiarowej uzyskała najgorsze rezultaty. Poszczególne nagrania tej osoby były mylone niemalże w każdym punkcie.



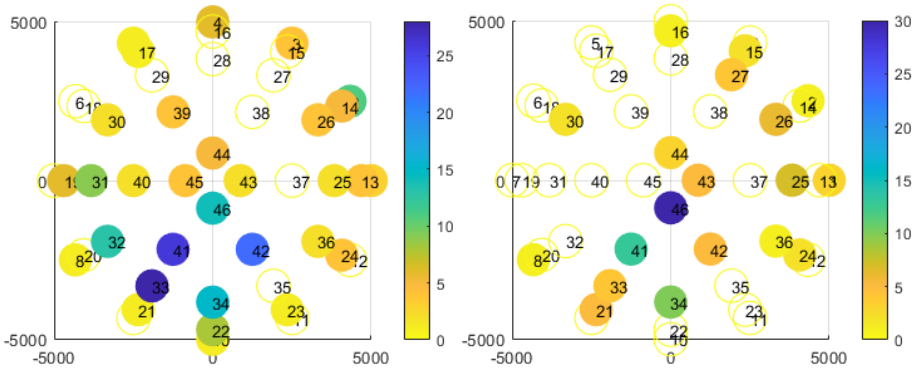
Rysunek 75: Kopuła wartości PBA sieci LSTM, dla akcji obroty, widziana z trzech różnych perspektyw (a) z boku, (b) z boku, (c) od góry.

Do pomyłek z akcją stanie dochodziło nie tylko w każdej z lokalizacji wirtualnej kamery, ale również każda z osób wykonująca to ćwiczenie została, dla co najmniej 4 różnych położeń nieprawidłowo zaklasyfikowana. Do największej liczby pomyłek z akcją stanie dochodziło,w punktach znajdujących się po lewej stronie pacjenta. Po przeprowadzeniu dokładniejszej analizy konkretnych nagrań danej osoby, które zostały błędnie sklasyfikowane wykazała pewne zależności pomiędzy poprawnością klasyfikacji a sposobem oraz kierunkiem obrotu. Obroty niezbyt dynamiczne były częściej klasyfikowane, jako stanie, gdy wirtualna kamera znajdowała się po skosie od pacjenta. Gdy znajdowała się na wprost/za pacjentem lub idealnie po jego prawej/lewej stronie, do pomyłek praktycznie nie dochodziło. Dodatkowo większość obrotów wykonywana była w prawą stronę, co może, ale nie musi mieć związku z tym, iż dla punktów znajdujących się po prawej stronie pacjenta osiągnięto nieco lepsze rezultaty.



(a) (b)

Rysunek 76: Kopuła liczby błędnych klasyfikacji wraz z ich numerami dla sieci LSTM, w których akcja obroty mylona była z akcjami: (a) schylanie się, (b) uderzenie



(a) (b)

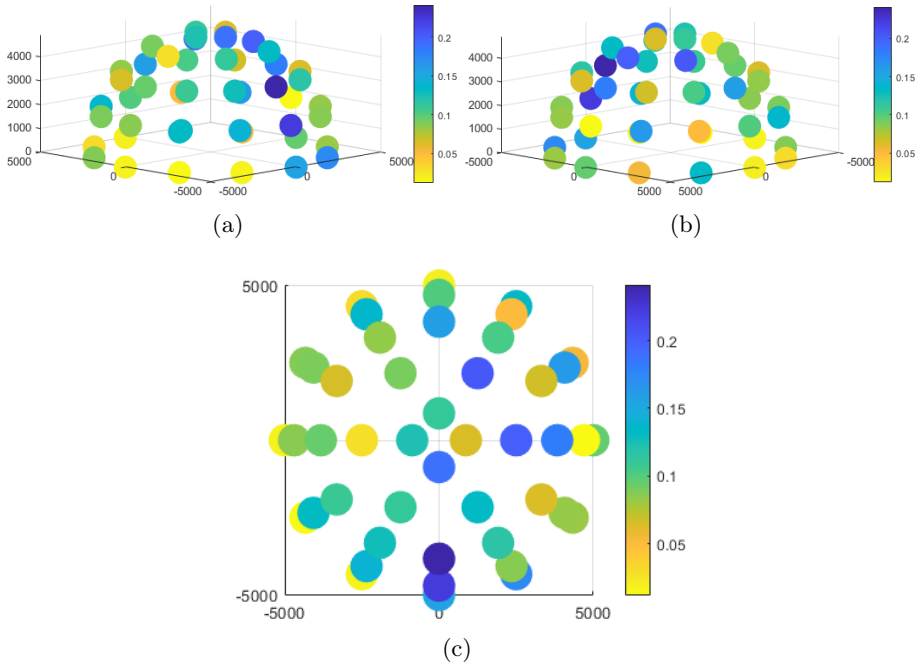
Rysunek 77: Kopuła liczby błędnych klasyfikacji wraz z ich numerami dla sieci LSTM, w których akcja obroty mylona była z akcjami: (a) kopnięcie wysokie proste, (b) kopnięcie wysokie boczne.

Tabela 11: Lista punktów, dla których dochodziło do pomyłek akcji obrót (na 28750 klasyfikacji) z innymi akcjami

Akcja	Punkty	Pomyłki
Chód osób zdrowych	11 23 27 37	5
Chód osób chorych	1 12 14 15 25 26 29 37 38 45	18
Stanie	Wszystkie	1176
Schylanie się	Wszystkie oprócz 3 4 5 8 9 12 13 17 ... 18 20 21 22 29 30 31 33 34 35 42	117
Uderzenie	Wszystkie oprócz 2 5 8 18 27 34 38 45	296
Kopnięcie niskie	1 2 13 14 19 24 25 30 31 33 42 44 45 46	29
Kopnięcie w. proste	Wszystkie oprócz 6 7 9 11 12 15 16 18... 20 27 28 29 35 37 38	211
Kopnięcie w. boczne	1 2 8 15 16 21 24 25 26 27 30 33 34 36 41 42 43 44 46 26	105

6.4.6 Akcja Schylanie się

Akcja schylanie się, spośród wszystkich akcji statycznych, cechowała się najgorszą rozpoznawalnością. W zależności od położenia wirtualnej kamery, czułość klasyfikacji tej akcji wynosiła od 75% do 98%. Wyższą czułość klasyfikacji, oraz wartość PBA uzyskiwano gdy wirtualna kamera znajdowała się bliżej podłoża (rys. 78).

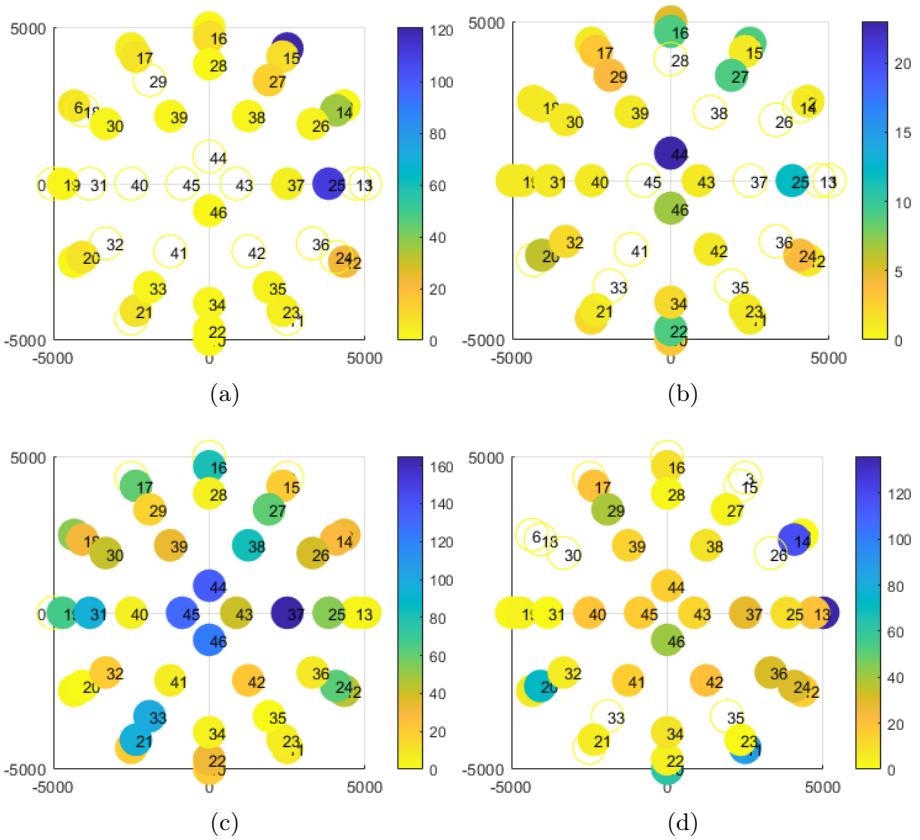


Rysunek 78: Kopuła wartości PBA sieci LSTM, dla akcji schylanie się, widziana z trzech różnych perspektyw (a) z boku, (b) z boku, (c) od góry.

W każdym z punktów dochodziło do pomyłek, z co najmniej dwoma innymi akcjami (tab. 12). Błędna klasyfikacja, jako jeden wariantów akcji chód zdarzała się w wielu punktach, jednakże pomyłki te były w większości sporadyczne. Najczęściej z danej perspektywy mylone były 1-2 nagrania danej osoby. Osoby te były różne w zależności od położenia wirtualnej kamery. W przypadku obu tych akcji znalazły się punkty, w których do pomyłek dochodziło znacznie częściej. W przypadku akcji chód osób zdrowych były to punkty znajdujące się z lewej strony osoby, oraz za osobą lekko z lewej - łącznie 120 pomyłek w każdym z tych punktów. W przypadku pomyłek z akcją chód osób zdrowych najczęściej do pomyłek dochodziło, gdy wirtualna kamera znajdowała się za osobą blisko szczytu kopuły.

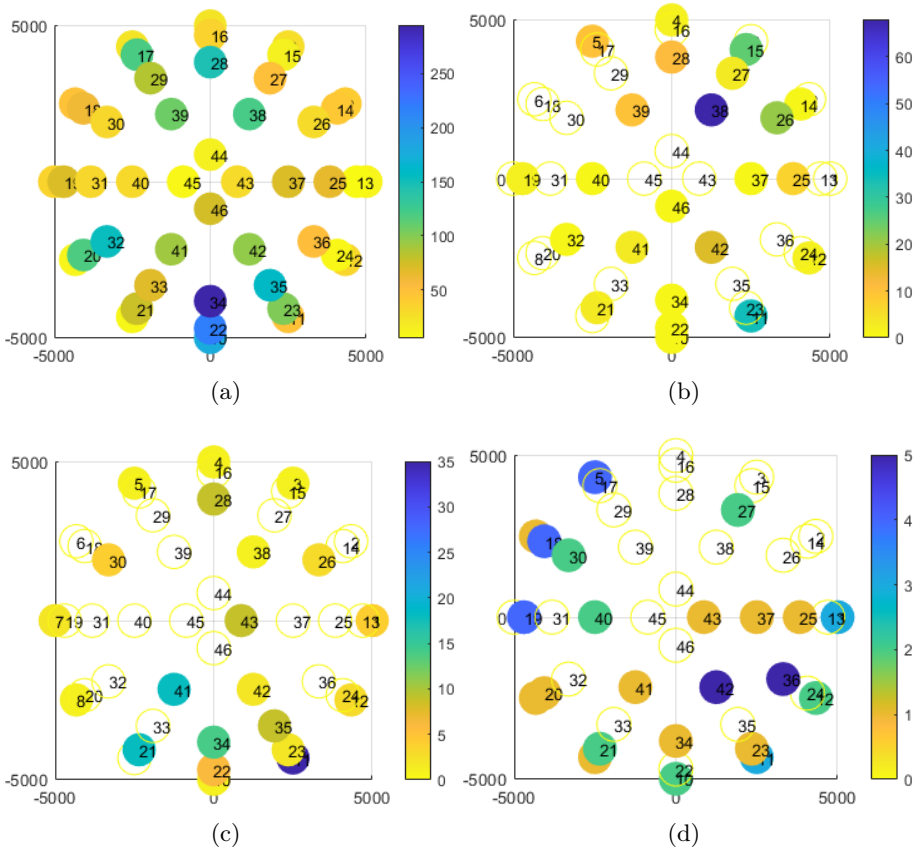
Tabela 12: Lista punktów, dla których dochodziło do pomyłek akcji schylanie się (na 60536 klasyfikacji) z innymi akcjami

Akcja	Punkty	Pomyłki
Chód osób zdrowych	2 3 4 5 6 8 10 12 14 15 16 17 19 20 21 22 23 25 26 27 28 30 33 34 35 37 38 39 46	390
Chód osób chorych	Wszystkie oprócz 1 8 13 14 26 28 33 35... 36 37 38 41 45	126
Stanie	Wszystkie oprócz 3 4 5 7	1856
Obroty	Wszystkie - oprócz 3 4 5 6 9 15 26 30 33 35	885
Uderzenie	Wszystkie	3153
Kopnięcie niskie	4 5 10 11 12 14 15 19 21 22 25 26 27 28 32 34 37 38 39 40 41 42 46	224
Kopnięcie w. proste	1 3 4 5 7 8 10 11 12 21 22 23 26 28 30 34 35 38 41 42 43	144
Kopnięcie w. boczne	1 5 6 8 9 10 11 12 18 19 20 21 23 25 27 30 34 36 37 40 41 42 43	50



Rysunek 79: Kopuła liczby błędnych klasyfikacji wraz z ich numerami dla sieci LSTM, w których akcja schylanie się mylona była z akcjami: (a) chód osób zdrowych, (b) chód osób chorych, (c) stanie, (d) obroty.

Z pozostałymi akcjami statycznymi dochodziło do znacznie większej liczby pomyłek. Ruch ponad połowy osób, co najmniej raz, z minimum jednej perspektywy był błędnie klasyfikowany, jako stanie. Sumarycznie do największej liczby pomyłek dochodziło w punktach znajdujących się nad osobą lub z lewej strony blisko szczytu kopuły - maksymalnie do 160 pomyłek na punkt. W przypadku akcji obroty, liczba osób, które zostały, co najmniej raz błędnie zaklasyfikowane, była nieco mniejsza. W przypadku obu tych akcji, gdy wirtualna kamera znajdowała się za osobą i blisko podłoża (punkty 9, 10 i 11) nie dochodziło do żadnych pomyłek. Natomiast dla punktów po lewej stronie osoby, bliżej podłoża dochodziło do większej liczby błędów - do 128 pomyłek w punkcie.



Rysunek 80: Kopuła liczby błędnych klasyfikacji wraz z ich numerami dla sieci LSTM, w których akcja schylanie się mylona była z akcjami: (a) uderzenie, (b) kopnięcie niskie, (c) kopnięcie wysokie proste, (d) kopnięcie wysokie boczne

W przypadku akcji uderzenie, do pomyłek dochodziło w każdym z położeń wirtualnej kamery. Błędy te dotyczyły 70% osób wykonujących tą akcję. Średnio 6 nagrań danej osoby było błędnie klasyfikowanych, jako uderzenie dla danej lokalizacji wirtualnej kamery. Przy czym gdy wirtualna kamera ustawiona była na wprost osoby, do pomyłek dochodziło znacznie częściej - do 280 pomyłek w danym punkcie. Cechą wspólną wszystkich nagrań, które zostały błędnie zaklasyfikowane są dodatkowe wymachy rąk podczas wykonywania skłonu.

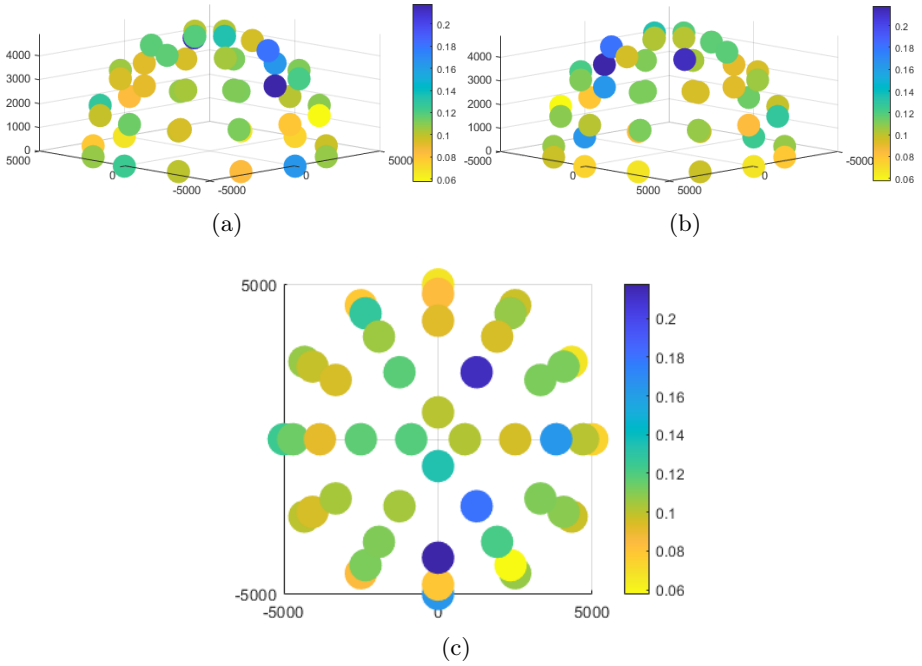
W przypadku wszystkich trzech rodzajów kopnięć do pomyłek dochodziło znacznie rzadziej. W większości punktów liczba pomyłek nie przekraczała 5 nagrań różnych osób. Ponownie punktami w których dochodziło do największej liczby błędów były punkty znajdujące się z lewej strony nagrywanej osoby. Dodatkowo w przypadku akcji kopnięcie niskie, łączna liczba pomyłek w punkcie była większa dla punktów znajdujących się za osobą.

Wśród wszystkich osób wykonujących skłon, znalazła się jedna osoba, której ruch w zależności od położenia wirtualnej kamery był klasyfikowany, jako każda z pozostałych akcji. Dane nagranie widziane z przodu było klasyfikowane, jako

chód osoby zdrowej, a z góry, jako stanie. Cechą wspólną wszystkich tych nagrań jest ich bardzo krótki czas trwania - mniej niż sekunda.

6.4.7 Akcja Uderzenie

Pierwsza z akcji niebezpiecznych - uderzenie, dla większości położeń wirtualnej kamery cechowała się dość dobrą rozpoznawalnością (rys. 81). Czułość klasyfikacji w najgorszych puntach była na poziomie 78%, w najlepszych 94%. Nieco słabsze wyniki, ponownie, osiągano, gdy kamera znajdowała się z lewej strony rejestrowanej osoby.

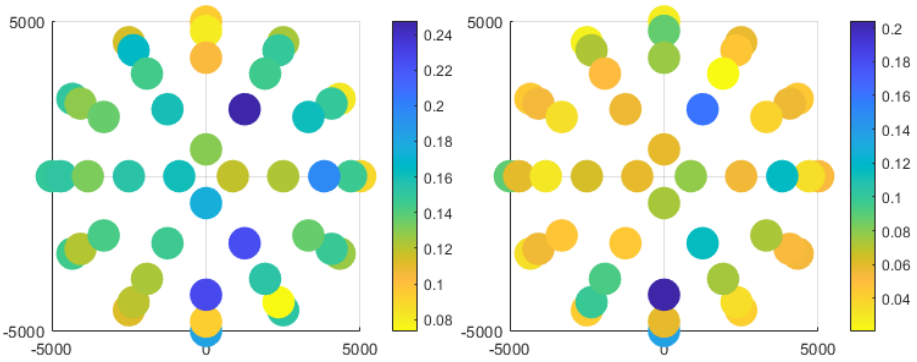


Rysunek 81: Kopuła wartości PBA sieci LSTM, dla akcji uderzenie, widziana z trzech różnych perspektyw (a) z boku, (b) z boku, (c) od góry.

Akcja ta była mylona nie tylko z pozostałymi akcjami niebezpiecznymi, ale i również akcjami statycznymi. Dodatkowo dla kilku położeń wirtualnej kamery (za/przed osobą, lub na szczycie kopuły) dochodziło do jednostkowych pomyłek z obiema akcjami chód. Podobnie jak w przypadku danych trójwymiarowych, szczegółowa analiza błędów została przeprowadzona osobno dla zawodników Karate i Taekwondo.

Na rysunku 82 przedstawiono kopuły wartości PBA, widziane z góry, dla zawodników obu tych dyscyplin. Podobnie jak w przypadku danych trójwymiarowych, zawodnicy Karate uzyskali gorsze rezultaty dla większości położeń wirtualnej kamery. Wyjątek stanowiły punkty znajdujące się za zawodnikiem. Zawodnicy Taekwondo osiągali nieco lepsze rezultaty, gdy wirtualna kamera znajdowała się z ich prawej strony. Zawodnicy Karate osiągali podobne rezultaty dla większości położeń wirtualnej kamery, za wyjątkiem wspomnianych trzech

punktów znajdujących się za zawodnikiem, oraz dwóch punktów na wprost.



(a) (b)

Rysunek 82: Kopuła wartości PBA sieci LSTM, dla akcji uderzenie, widziana z góry dla zawodników (a) Karate, (b) Taekwondo.

W tabelach 13 i 14 przedstawiono listy punktów, w których dochodziło do pomyłek z daną klasą wraz z informacją o łącznej liczbie pomyłek we wszystkich wymienionych punktach, kolejno dla zawodników Karate oraz Taekwondo. W przypadku zawodników obu dyscyplin dochodziło do pomyłek ze wszystkimi pozostałymi klasami. Przy czym do pomyłek z akcjami chód dochodziło tylko w kilku położeniach wirtualnej kamery (głównie, gdy kamera była na wprost zawodnika) i tylko w przypadku jednego bądź dwóch powtórzeń uderzenia, wykonywanego przez jedną osobę, inną w każdym z punktów.

Tabela 13: Lista punktów, dla których dochodziło do pomyłek akcji uderzenie (na 27692 klasyfikacji) z innymi akcjami, wśród zawodników Karate

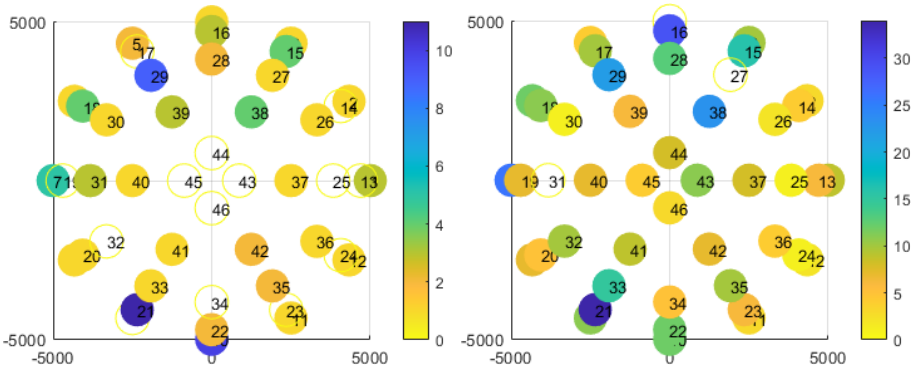
Akcja	Punkty	Pomyłki
Chód osób zdrowych	10 33 35 38 45 46	14
Chód osób chorych	2 14 31 37 38 44 45	7
Stanie	Wszystkie oprócz 9 14 17 19 24 25 32 34... 43 44 45 46	85
Obroty	Wszystkie oprócz 1 8 16 23 25 27	331
Schylanie się	Wszystkie oprócz 15 17	290
Kopnięcie niskie	Wszystkie	781
Kopnięcie w. proste	Wszystkie	1603
Kopnięcie w. boczne	Wszystkie oprócz 23	825

Do pomyłek z akcją stanie dużo częściej dochodziło w przypadku zawodników Taekwondo. Znacznie częściej, gdy wirtualna kamera znajdowała się po prawej stronie zawodnika (rys. 83). Co ciekawe, do największej liczby pomyłek z tą akcją doszło w punkcie znajdującym się na wprost zawodnika lekko z prawej (łącznie 34 pomyłki). W przypadku zawodników Karate również do większej liczby pomyłek dochodziło, gdy wirtualna kamera znajdowała się bar-

dziej z prawej strony zawodnika lub tuż za nim. Dodatkowo punkt, w którym doszło do największej sumarycznej liczby błędów (34 błędy) był tym samym, co w przypadku zawodników Taekwondo. Fakt, iż do większej liczby pomyłek dochodziło wśród zawodników Taekwondo, może mieć związek z dynamiką wykonywanego uderzenia. Zawodnicy Taekwondo wykonywali znacznie szybsze, a tym samym trwające krócej uderzenia. Zazwyczaj też stali w miejscu i nie wykonywali dodatkowych ruchów nogami. W przypadku zawodników Karate, punkty, w których nie dochodziło do żadnych błędów z omawianą klasą znajdowały się przed nagrywaną osobą.

Tabela 14: Lista punktów, dla których dochodziło do pomyłek akcji uderzenie (na 32338 klasyfikacji) z innymi akcjami, wśród zawodników Taekwondo

Akcja	Punkty	Pomyłki
Chód osób zdrowych	35	3
Chód osób chorych	14 16 21 33 35 37	7
Stanie	Wszystkie oprócz 4 31	414
Obroty	Wszystkie oprócz 4 6 13 26 27 31	280
Schylanie się	Wszystkie oprócz 4 13 15 23 27	411
Kopnięcie niskie	Wszystkie oprócz 1 3 5 6 7 8 9 11 14 16 22 23 27 28 29 30 35 40 41 46	142
Kopnięcie w. proste	Wszystkie oprócz 3 5 6 7 8 9 11 12 13 14 16 19 20 21 23 27 28 29 30 31 39	323
Kopnięcie w. boczne	1 11 21 25 26 33 41	10

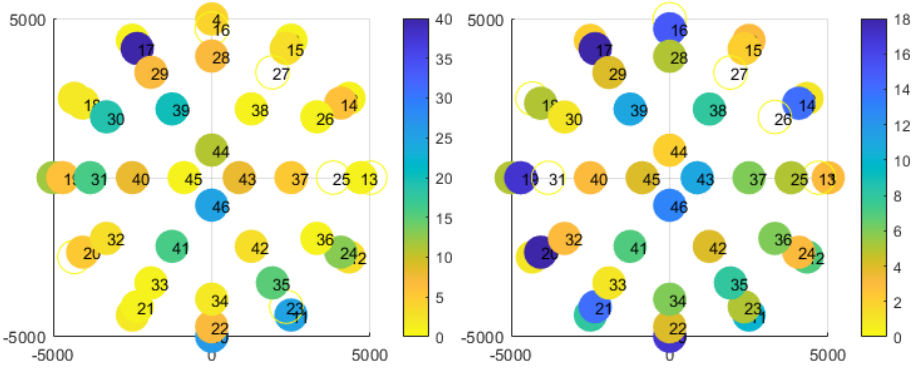


(a) (b)

Rysunek 83: Kopuła liczby błędnych klasyfikacji wraz z ich numerami dla sieci LSTM, w których akcja uderzenie mylona była z akcją stanie dla zawodników (a) Karate, (b) Taekwondo.

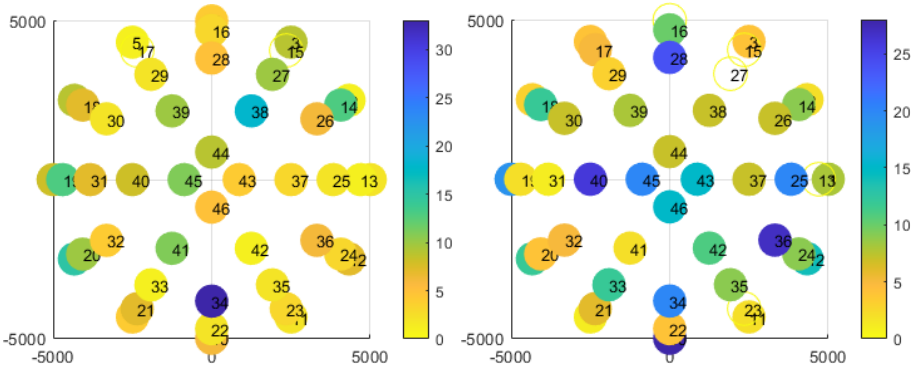
W przypadku pozostałych dwóch akcji statycznych, do pomyłek wśród zawodników Karate dochodziło w znacznie większej liczbie położeń wirtualnej kamery. Do pomyłek z akcją obroty częściej dochodziło wśród zawodników Karate, a z akcją schylanie się wśród zawodników Taekwondo. W przypadku obu grup,

do błędnej klasyfikacji nie dochodziło, gdy wirtualna kamera znajdowała się dość nisko za zawodnikiem (rys. 84 i 85). Dodatkowo rozkład łącznej liczby pomyłek na punkt wśród zawodników Karate jest w miarę równomierny, podczas gdy nagrania zawodników Taekwondo częściej są mylone, gdy wirtualna kamera znajduje się z prawej strony zawodnika.



(a) (b)

Rysunek 84: Kopuła liczby błędnych klasyfikacji wraz z ich numerami dla sieci LSTM, w których akcja uderzenie mylona była z akcją obroty dla zawodników (a) Karate, (b) Taekwondo.



(a) (b)

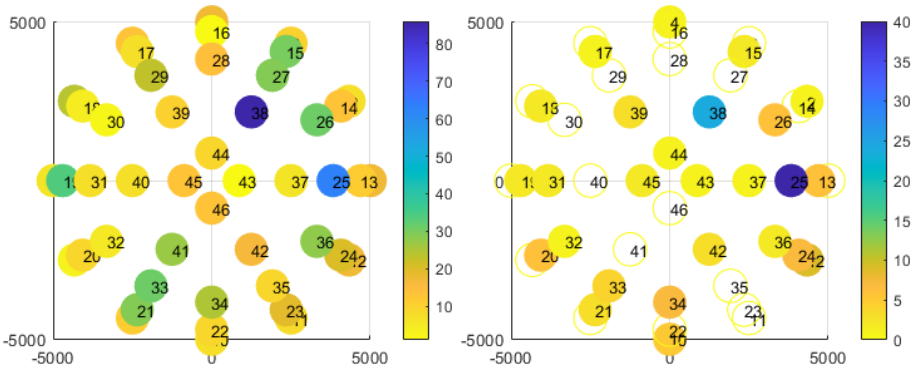
Rysunek 85: Kopuła liczby błędnych klasyfikacji wraz z ich numerami dla sieci LSTM, w których akcja uderzenie mylona była z akcją schylanie się dla zawodników (a) Karate, (b) Taekwondo.

Do pomyłek z akcją kopnięcie niskie dochodziło przede wszystkim wśród zawodników Karate. Znacznie częściej, gdy wirtualna kamera znajdowała się po lewej stronie zawodnika, przy czym ich ilośc malała bliżej podłoża było (rys. 86 a). W przypadku zawodników Taekwondo, jeśli dochodziło do pomyłek to

głównie dla średniej wysokości wirtualnej kamery, oraz dla punktów za lub po lewej stronie zawodnika (rys. 86 b). Dla pozostałych położeń wirtualnej kamery liczba błędów nie przekraczała 10 na punkt.

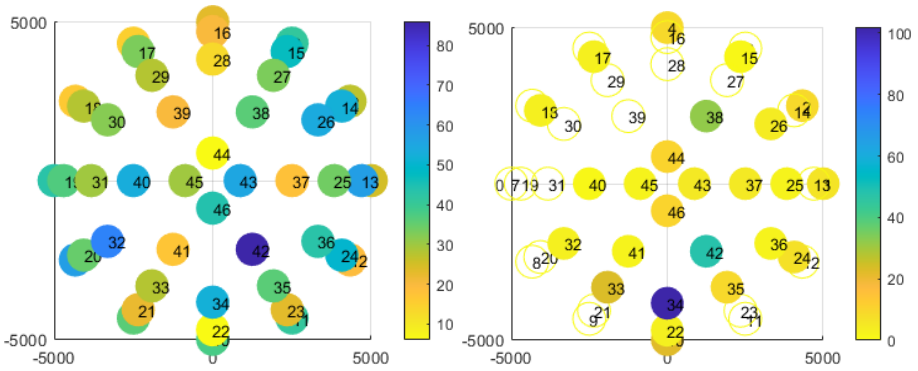
Podobnie w przypadku obu rodzajów kopnięć wysokich, uderzenia zawodników karate były mylone z kopnięciami wysokimi praktycznie dla wszystkich położeń wirtualnej kamery. Z kopnięciem prostym mniej, gdy wirtualna kamera znajdowała się za zawodnikiem, a z bocznym, gdy za, z lewej strony (rys. 87 a i 88 a). Uderzenia zawodników Taekwondo częściej były mylone z kopnięciem wysokim prostym. Większość punktów, dla których dochodziło do pomyłek znajdowała się na wprost lub z lewej strony zawodnika (rys. 87 b). Do pomyłek z kopnięciem wysokim bocznym dochodziło w sporadycznie (rys. 88 b).

W przypadku zawodników obu sportów walki, ponownie można zauważyć wpływ rodzaju uderzenia na błędy w klasyfikacji. Zdecydowana większość pomyłek dotyczyła nagrań, w których zawodnik uderzał w tarczę.



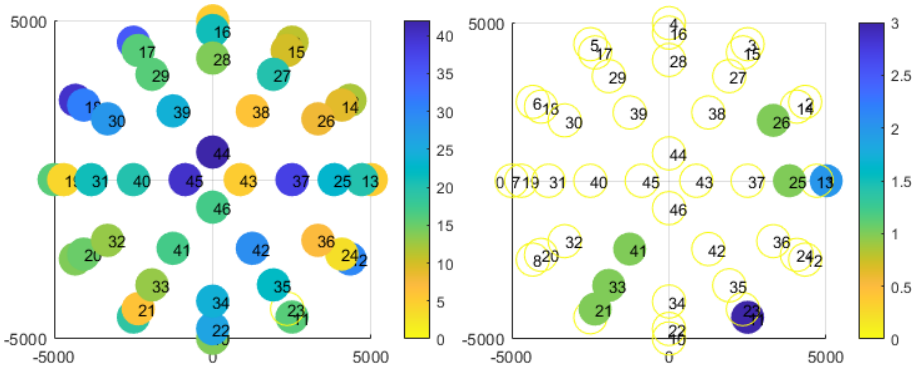
(a) (b)

Rysunek 86: Kopuła liczby błędnych klasyfikacji wraz z ich numerami dla sieci LSTM, w których akcja uderzenie mylona była z akcją kopnięcie niskie dla zawodników (a) Karate, (b) Taekwondo.



(a) (b)

Rysunek 87: Kopuła liczby błędnych klasyfikacji wraz z ich numerami dla sieci LSTM, w których akcja uderzenie mylona była z akcją kopnięcie wysokie proste dla zawodników (a) Karate, (b) Taekwondo.

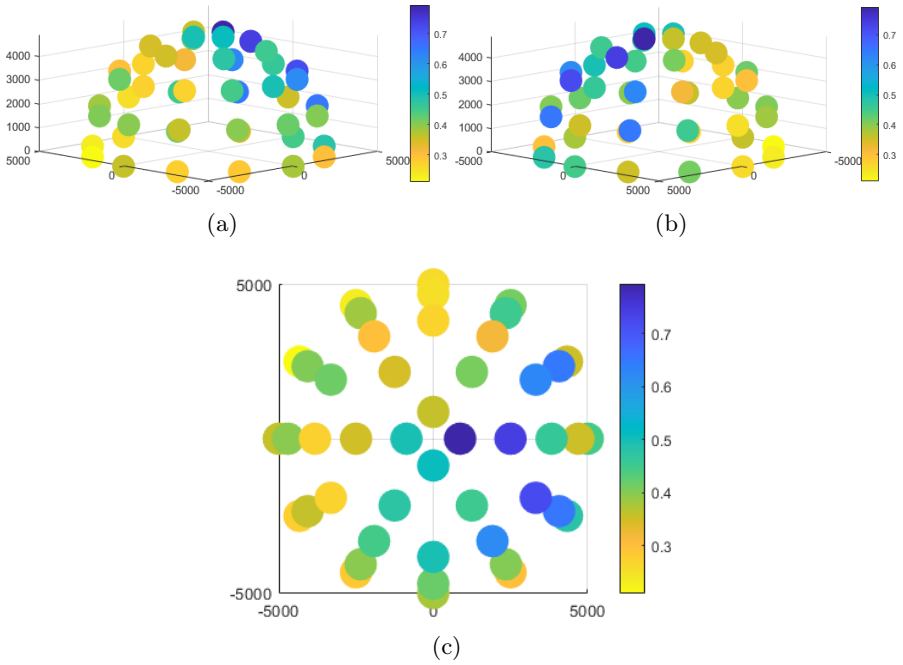


(a) (b)

Rysunek 88: Kopuła liczby błędnych klasyfikacji wraz z ich numerami dla sieci LSTM, w których akcja uderzenie mylona była z akcją kopnięcie wysokie boczne dla zawodników (a) Karate, (b) Taekwondo.

6.4.8 Akcja Kopnięcie niskie

Jakość klasyfikacji akcji kopnięcie niskie, w bardzo dużym stopniu zależała od położenia wirtualnej kamery. Dla punktów znajdujących się po lewej stronie zawodnika dochodziło do znacznie większej liczby pomyłek (rys. 89), a czułość klasyfikacji w najgorszym punkcie wynosiła tylko 20%. Punkty znajdujące się na wprost, za oraz po prawej stronie zawodnika uzyskiwały znacznie lepsze rezultaty, sięgające do 79%, co i tak w porównaniu do pozostałych akcji jest dość niskim wynikiem.



Rysunek 89: Kopuła wartości PBA sieci LSTM, dla akcji kopnięcie niskie, widziana z trzech różnych perspektyw (a) z boku, (b) z boku, (c) od góry.

Akcja ta mylona była ze wszystkimi pozostałymi akcjami. W tabeli 15 przedstawiono zestawienie punktów, w których dochodziło do pomyłek wraz z ich sumaryczną liczbą. Pomyłki z akcją chód osób zdrowych oraz chód osób chorych zdarzały się sporadycznie - jeśli już w jakimś punkcie dochodziło do pomyłki, dotyczyła ona najczęściej nie więcej jak 3 kopnięcia wykonane przez jednego lub dwóch zawodników. Wyjątek stanowił punkt, 22 w którym łącznie 10 kopnięć wykonanych przez 7 osób zostało błędnie sklasyfikowanych, jako chód osób chorych. Błędy te w dalszym ciągu pozostają jednak marginalne.

Podobnie w przypadku akcji stanie - do błędów dochodziło sporadycznie, w przypadku pojedynczych mniej dynamicznych kopnięć. Punkt, w którym dochodziło do największej liczby błędnych klasyfikacji (7 błędów) znajdował się za zawodnikiem, tuż nad podłożem.

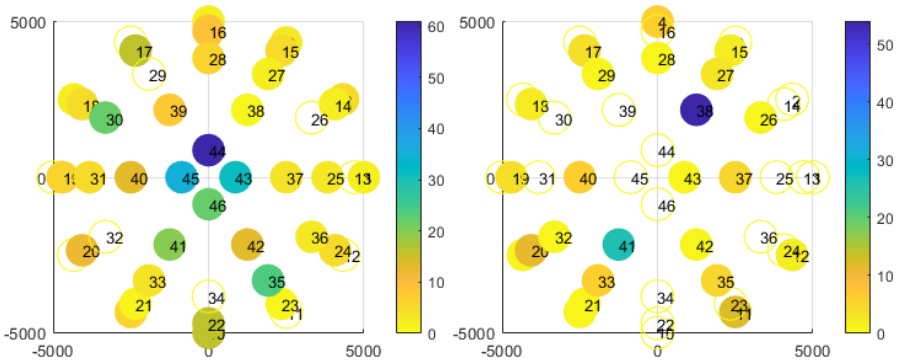
Tabela 15: Lista punktów, dla których dochodziło do pomyłek akcji kopnięcie niskie (na 49818 klasyfikacji) z innymi akcjami

Akcja	Punkty	Pomyłki
Chód osób zdrowych	3 7 27 29 30 33 34 35	21
Chód osób chorych	20 22 27 31 33 35 38 40 42	22
Stanie	1 4 7 12 19 24 28 30 31 34 37 42	30
Obroty	Wszystkie oprócz 5 7 8 11 12 13 26 29 32 34	391
Schylanie się	4 8 9 11 12 15 17 18 19 20 21 26 27 28 29 32 33 35 37 38 40 41 42 43	157
Uderzenie	Wszystkie	1680
Kopnięcie w. proste	Wszystkie	14881
Kopnięcie w. boczne	Wszystkie	3093

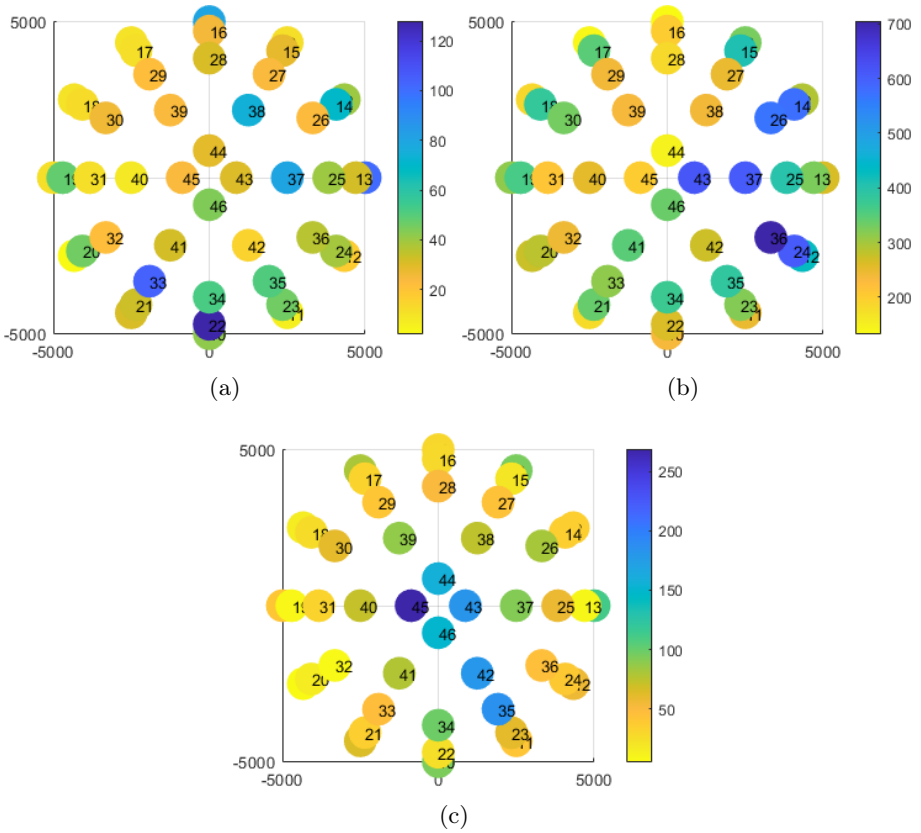
Do pomyłek z pozostałymi akcjami statycznymi dochodziło w większej liczbie punktów. W przypadku akcji obroty do największej liczby błędów dochodziło we wszystkich 4 punktach znajdujących się nad zawodnikiem (rys. 90 a). Wśród tych 4 punktów, najgorsze rezultaty uzyskał punkt 44, znajdujący się za zawodnikiem - łącznie 60 błędnie sklasyfikowanych kopnięć. Pomyłki z akcją schylanie się zdarzały się w większości punktów sporadycznie (rys. 90 b), za wyjątkiem punktu 38, znajdującego się z lewej strony za zawodnikiem, tuż przed szczytem kopuły. W punkcie tym doszło łącznie do 53 pomyłek. W przypadku obu omawianych akcji, pomyłki dotyczyły kilku zawodników, każdego w 5-6 różnych punktach.

Do pomyłek z pozostałymi akcjami niebezpiecznymi dochodziło dla wszystkich położeń wirtualnej kamery. Do pomyłek z akcją uderzenie rzadziej dochodziło, gdy kamera znajdowała się po prawej stronie za zawodnikiem (rys. 91 a). Kopnięcie każdego z zawodników było pomylone, co najmniej raz dla jednego z położeń wirtualnej kamery. Średnio, co najmniej jedno kopnięcie danego zawodnika było błędnie klasyfikowane, jako uderzenie w 13 punktach, przy czym znalazły się osoby, których uderzenia były mylone w ponad 20 punktach. Przy czym punkty te były różne dla każdej osoby.

Akcja uderzenie wysokie proste była akcją, z która doszło do największej łącznej liczby pomyłek. Każdy z zawodników został błędnie zaklasyfikowany, chociaż raz średnio w 35 punktach, minimalnie w 13, maksymalnie we wszystkich 46 (jeden zawodnik). Ponownie, sumarycznie, znacznie więcej błędów było, gdy wirtualna kamera znajdowała się po lewej stronie zawodnika (rys. 91 b). Analiza nagrań, które były błędnie klasyfikowane wykazała podobną zależność jak w przypadku danych trójwymiarowych - błędy dotyczyły kopnięć w powietrze.



Rysunek 90: Kopuła liczby błędnych klasyfikacji wraz z ich numerami dla sieci LSTM, w których akcja kopnięcie niskie mylona była z akcjami: (a) obroty, (b) schylanie się.



(a) (b)

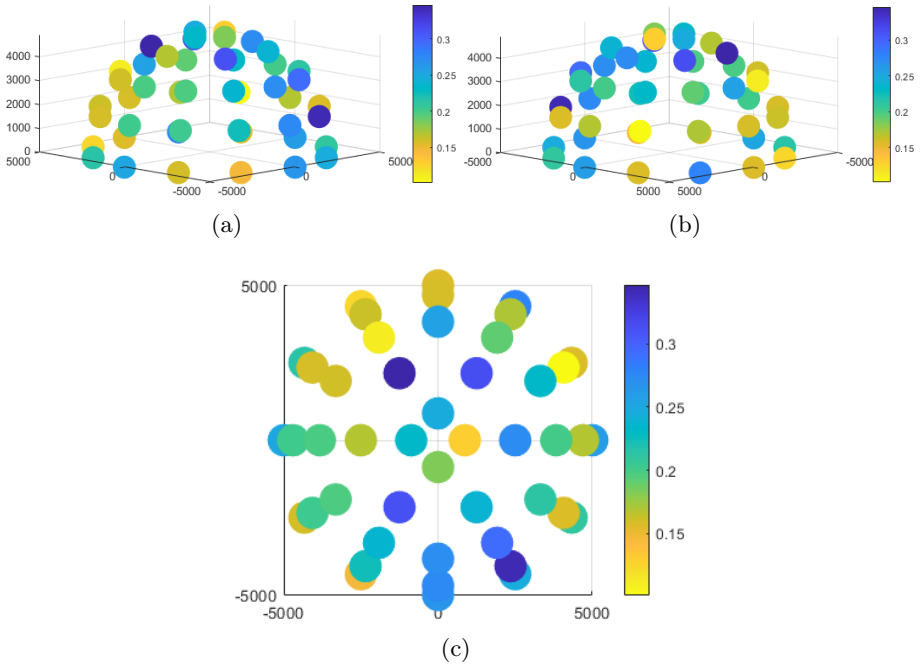
(c)

Rysunek 91: Kopuła liczby błędnych klasyfikacji wraz z ich numerami dla sieci LSTM, w których akcja kopnięcie niskie mylona była z akcjami: (a) uderzenie, (b) kopnięcie wysokie proste, (c) kopnięcie wysokie boczne.

Do pomyłek z akcją kopnięcie wysokie boczne dochodziło znacznie mniej, głównie, gdy wirtualna kamera znajdowała się na szczycie kopuły (rys. 91 c). Każdy z zawodników mylony były z tą akcją, średnio dla 15 położeń wirtualnej kamery. Do błędów znacznie rzadziej dochodziło, gdy kamera znajdowała się po prawej stronie zawodnika, lub była dość blisko podłoża.

6.4.9 Akcja kopnięcie wysokie proste

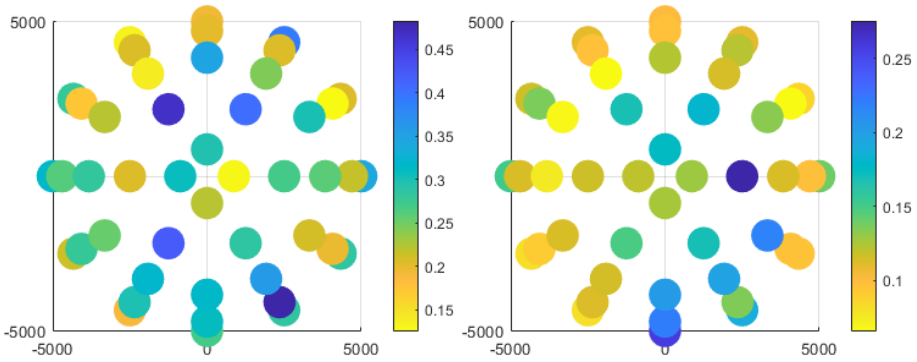
Rozpoznawalność akcji kopnięcie wysokie boczne również w znacznym stopniu zależała od położenia wirtualnej kamery. Czułość klasyfikacji tej akcji wahała się od 65% do 90%. Najlepsze rezultaty uzyskano dla punktów znajdujących się za zawodnikami lekko z lewej i z prawej strony (rys. 92). Akcja ta mylona była najczęściej z pozostałymi akcjami niebezpiecznymi dla każdego położenia wirtualnej kamery. Największa liczba pomyłek dotyczyła akcji kopnięcie niskie.



Rysunek 92: Kopuła wartości PBA sieci LSTM, dla akcji kopnięcie wysokie proste, widziana z trzech różnych perspektyw (a) z boku, (b) z boku, (c) od góry.

Ponieważ akcja ta została wykonana przez zawodników obu dyscyplin sportowych, dalsza bardziej szczegółowa analiza zostanie przeprowadzona osobno dla zawodników Karate i Taekwondo. Na rysunku 93 przedstawiono kopuły PBA widziane z góry dla zawodników obu dyscyplin. Dodatkowo w tabelach 16 i 17 przedstawiono listę punktów, w których dochodziło do pomyłek z daną akcją kolejno dla zawodników Karate i Taekwondo.

Ponownie nagrania zawodników Karate cechowały się znacznie gorszą, jakością klasyfikacji. Najlepsze rezultaty osiągali oni, gdy wirtualna kamera znajdowała się za zawodnikiem, dość blisko podłoża. Kopnięcia zawodników Taekwondo były natomiast znacznie rzadziej mylone. Pomyłki głównie zdarzały się, gdy wirtualna kamera znajdowała się w jednym z położeń znajdujących się między na wprost zawodnika a jego lewą stroną.



(a) (b)

Rysunek 93: Kopuła wartości PBA sieci LSTM, dla akcji kopnięcie wysokie proste, widziana z góry dla zawodników (a) Karate, (b) Taekwondo.

Tabela 16: Lista punktów, dla których dochodziło do pomyłek akcji kopnięcie wysokie proste (na 29532 klasyfikacji), wśród zawodników Karate

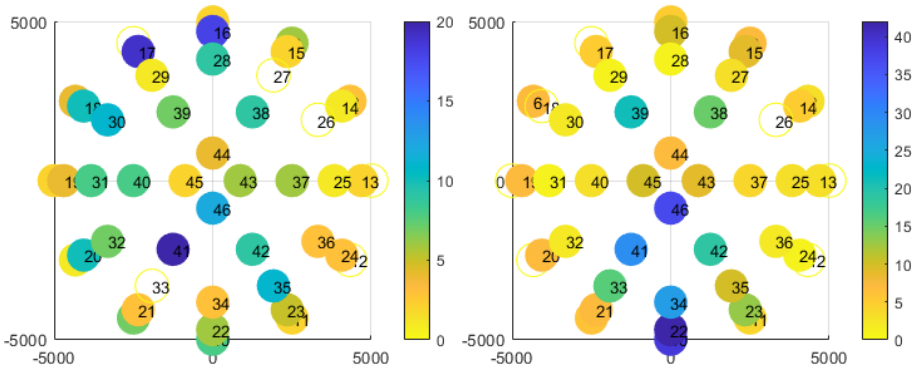
Akcja	Punkty	Pomyłki
Chód osób zdrowych	22 27 33 35 37 45	14
Chód osób chorych	11 14 15 17 21 22 27 37 38 40 42 45	19
Stanie	4 8 10 23	9
Obroty	Wszystkie oprócz 1 5 12 26 27 33	255
Schylanie się	10 12 16 20 21 22 23	41
	25 29 34 35 38 41 46	
Uderzenie	Wszystkie	1211
Kopnięcie niskie	Wszystkie	4046
Kopnięcie w. boczne	Wszystkie	2340

Tabela 17: Lista punktów, dla których dochodziło do pomyłek akcji kopnięcie wysokie proste (na 62652 klasyfikacji) z innymi akcjami, wśród zawodników Taekwondo

Akcja	Punkty	Pomyłki
Chód osób zdrowych	1 2 3 10 27 29 33 35 36 37 40 46	27
Chód osób chorych	2 3 11 12 14 22 27 34 35 37 38 40	36
Stanie	1 4 9 10 15 21 22 23 24 25 33 34 35 36 37 41 42 43	65
Obroty	Wszystkie oprócz 1 5 7 8 12 18 26	402
Schylanie się	4 8 12 19 22 29 34 35 36 37 38 39 41 42 43 46	54
Uderzenie	Wszystkie oprócz 5 7 8 19 20 29 30 40	650
Kopnięcie niskie	Wszystkie	1150
Kopnięcie w. boczne	Wszystkie	4244

W przypadku zawodników obu dyscyplin do pomyłek z akcjami chód osób zdrowych, chód osób chorych, stanie, oraz schylanie się dochodziło sporadycznie. Nieco częściej w przypadku zawodników Taekwondo, dla punktów znajdujących się na wprost lub lekko z lewej strony zawodnika. Pomyłki te dotyczyły najczęściej różnych osób w różnych punktach, najczęściej jednego/dwóch kopnięć danej osoby.

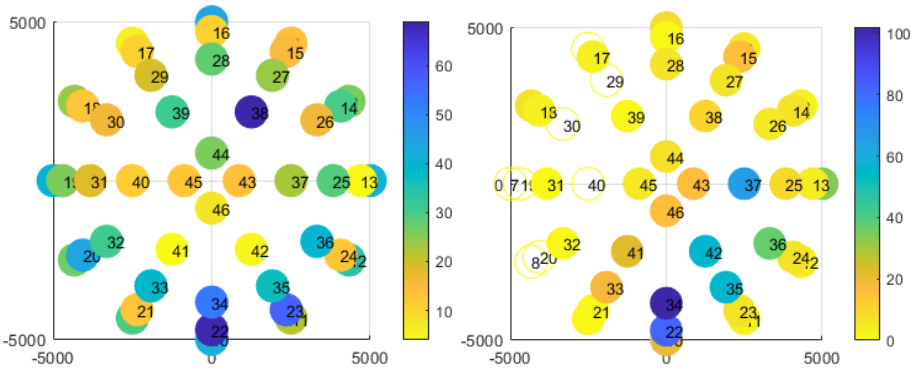
Nieco częściej, wśród zawodników obu dyscyplin dochodziło do pomyłki z akcją obroty. Na rysunku 94 przedstawiono kopuły punktów wraz z numerami, kolor danego punktu zależy od łącznej liczby pomyłek. Kopnięcia wysokie zawodników Taekwondo były znacznie częściej mylone z obrotami, gdy wirtualna kamera znajdowała się przed zawodnikiem. Punkty znajdujące się idealnie na wprost zawodnika uzyskały najgorsze rezultaty - do 42 błędnych klasyfikacji. Pomyłki te dotyczyły wszystkich zawodników, z czego najczęściej każdy z nich był mylony w 1-2 różnych punktach. W przypadku zawodników Karate, błędy dotyczyły tylko połowy z nich. Średnio pomyłki dla jednej z mylonych osób zdarzały się dla jednego jej nagrania widzianego z 6 różnych perspektyw. Do pomyłek znacznie rzadziej dochodziło, gdy wirtualna kamera znajdowała się z lewej strony zawodnika.



(a) (b)

Rysunek 94: Kopuła liczby błędnych klasyfikacji wraz z ich numerami dla sieci LSTM, w których akcja kopnięcie wysokie proste mylona była z akcją obroty dla zawodników: (a) Karate, (b) Taekwondo.

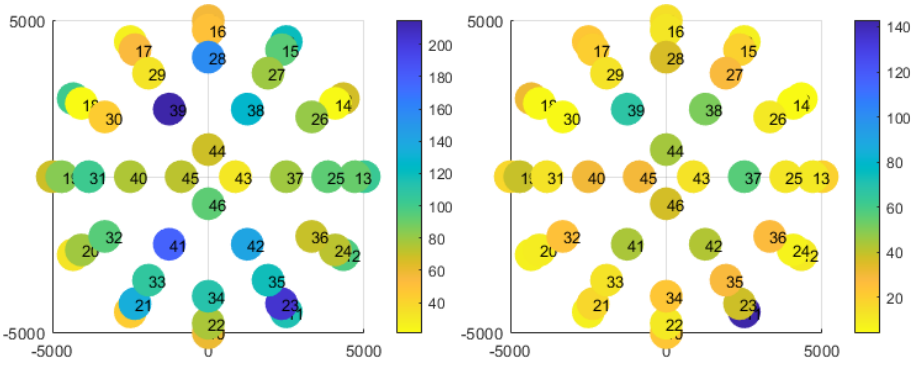
Do pomyłek z akcją uderzenie dochodziło znacznie częściej. Kopnięcia zawodników Karate były znacznie częściej błędnie klasyfikowane, jako uderzenie, gdy wirtualna kamera znajdowała się przed zawodnikiem, lub dość blisko podłoża (rys. 95 a). Pomyłki dotyczyły praktycznie wszystkich zawodników. Średnio jedna osoba była mylona z tą akcją dla 13 różnych położeń wirtualnej kamery. Co najmniej jedno z wysokich kopnięć każdego z zawodników Taekwondo było błędnie klasyfikowanych, jako obrót. Sytuacja ta najczęściej miała miejsce, gdy wirtualna kamera znajdowała się przed zawodnikiem. Gdy ustawiona była z jego prawej strony do błędów dochodziło rzadziej lub wcale.



(a) (b)

Rysunek 95: Kopuła liczby błędnych klasyfikacji wraz z ich numerami dla sieci LSTM, w których akcja kopnięcie wysokie proste mylona była z akcją uderzenie dla zawodników: (a) Karate, (b) Taekwondo.

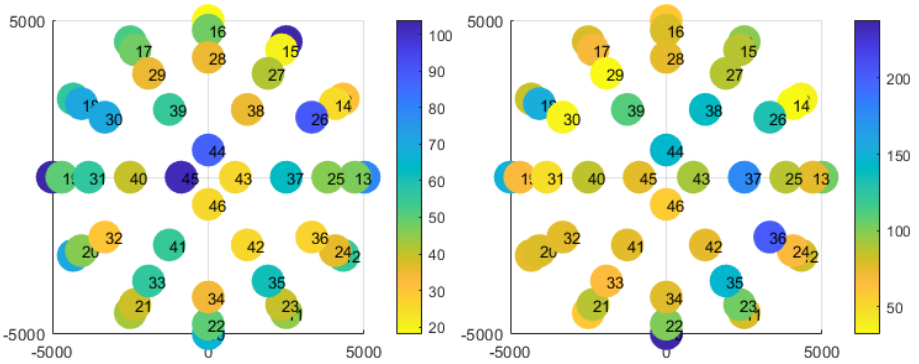
Największa liczba wysokich kopnięć zawodników Karate mylona była z kopnięciem niskim. Co najmniej jedno nagranie danego zawodnika było mylone, w co najmniej jednym z punktów. Średnio dochodziło do pomyłek w 21 punktach dla danej osoby. Znalazł się też zawodnik, którego uderzenia były mylone w praktycznie każdym położeniu wirtualnej kamery. Do pomyłek w tej grupie dochodziło trochę rzadziej, gdy kamera znajdowała się za i lekko z prawej strony zawodnika (rys. 96 a). Kopnięcia zawodników Taekwondo były znacznie mniej mylone z tą akcją. Do pomyłek wśród zawodników Taekwondo dochodziło znacznie rzadziej. Najczęściej, gdy wirtualna kamera była dość blisko szczytu kopuły (rys. 96 b). Pomyłki dotyczyły praktycznie wszystkich zawodników, przy czym kopnięcia danego zawodnika były średnio mylone w 13 punktach.



(a) (b)

Rysunek 96: Kopuła liczby błędnych klasyfikacji wraz z ich numerami dla sieci LSTM, w których akcja kopnięcie wysokie proste mylona była z akcją kopnięcie niskie dla zawodników: (a) Karate, (b) Taekwondo.

Do pomyłek z kopnięciem wysokim bocznym wśród zawodników Karate dochodziło rzadziej niż z kopnięciem niskim, ale częściej niż z uderzeniem. Do pomyłek dochodziło niezależne od położenia wirtualnej kamery (rys. 97 a). Dotyczyły one wszystkich zawodników - co najmniej jedno kopnięcie danego zawodnika było błędnie klasyfikowane, jako kopnięcie wysokie boczne dla średnio 17 różnych położeń wirtualnej kamery. W przypadku zawodników Taekwondo, ich wysokie kopnięcia proste były najczęściej mylone z bocznymi. Wykonanie tej techniki przez danego zawodnika było błędnie klasyfikowane dla średnio 23 położeń wirtualnej kamery. Do największej liczby pomyłek doszło, gdy wirtualna kamera znajdowała się na wprost zawodnika blisko podłoża (rys. 97 b). W punkcie tym doszło do 250 pomyłek nagrań, łącznie 14 różnych zawodników.



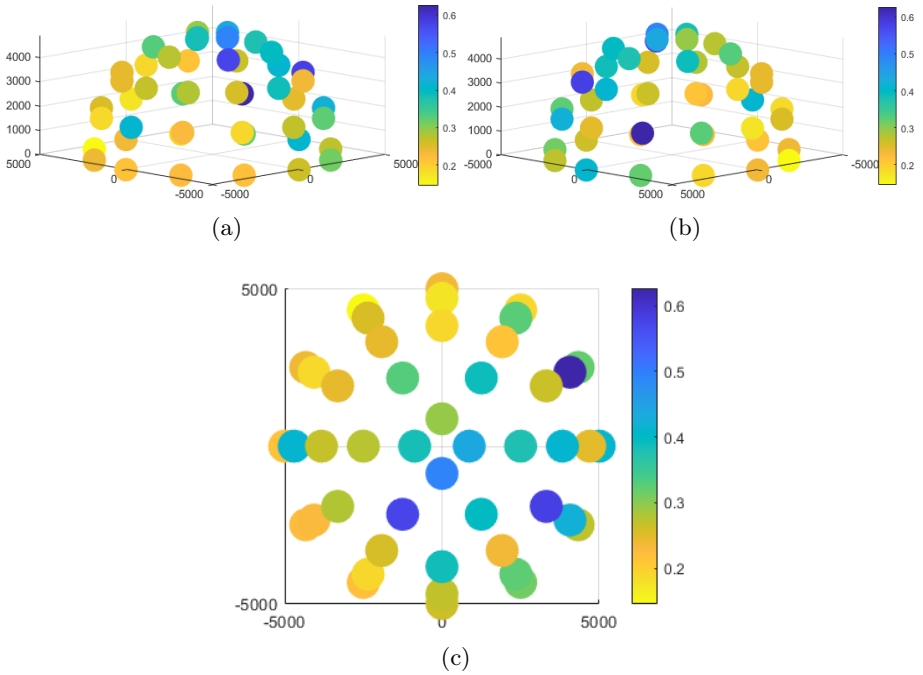
(a) (b)

Rysunek 97: Kopuła liczby błędnych klasyfikacji wraz z ich numerami dla sieci LSTM, w których akcja kopnięcie wysokie proste mylona była z akcją kopnięcie wysokie boczne dla zawodników: (a) Karate, (b) Taekwondo.

6.4.10 Akcja kopnięcie wysokie boczne

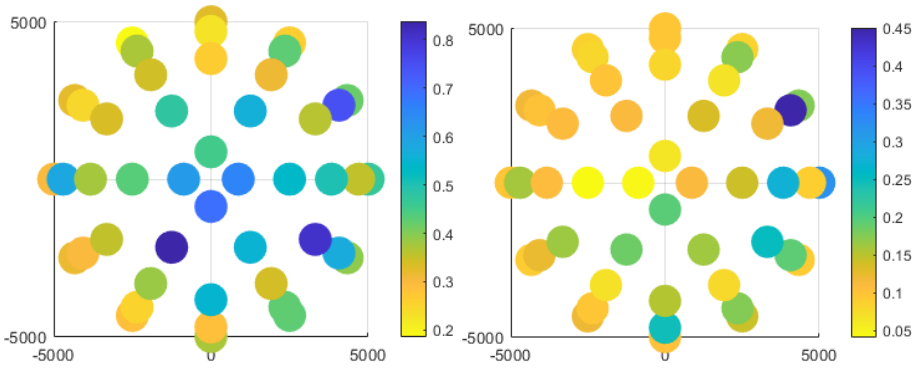
Ostatnią z omawianych akcji jest drugi wariant kopnięcia wysokiego - kopnięcie wysokie boczne. Czułość klasyfikacji tej akcji, oraz wartości PBA w bardzo dużym stopniu zależały od położenia wirtualnej kamery (rys. 98). Dla punktów znajdujących się po prawej stronie zawodnika, czułość klasyfikacji była lepsza i wynosiła do 85%, natomiast punkty znajdujące się po jego lewej stronie zazwyczaj miały znacznie gorszą, czułość wynoszącą w najgorszym punkcie tylko 37%.

Do pomyłek najczęściej dochodziło z drugim rodzajem wysokiego kopnięcia - kopnięciem wysokim prostym. Zdarzały się również pomyłki z każdą z pozostałych akcji. Ponieważ ten rodzaj kopnięcia również wykonywany był przez zawodników zarówno Karate jak i Taekwondo, dalsza bardziej szczegółowa analiza błędów została przeprowadzona z podziałem na te dyscypliny.



Rysunek 98: Kopuła wartości PBA sieci LSTM, dla akcji kopnięcie wysokie boczne, widziana z trzech różnych perspektyw (a) z boku, (b) z boku, (c) od góry.

Tak jak w przypadku pozostałych akcji niebezpiecznych, akcja ta znacznie częściej mylona była wśród zawodników Karate. Na rysunku 99 przedstawiono kopułę wartości PBA dla zawodników Karate oraz Taekwondo. Dodatkowo w tabelach 18 i 19 przedstawiono listę punktów, w których dochodziło do pomyłek wraz z łączną liczbą pomyłek w tych punktach.



(a) (b)

Rysunek 99: Kopuła wartości PBA sieci LSTM, dla akcji kopnięcie wysokie boczne, widziana z góry dla zawodników (a) Karate, (b) Taekwondo.

Tabela 18: Lista punktów, dla których dochodziło do pomyłek akcji kopnięcie wysokie proste (na 29256 klasyfikacji), wśród zawodników Karate

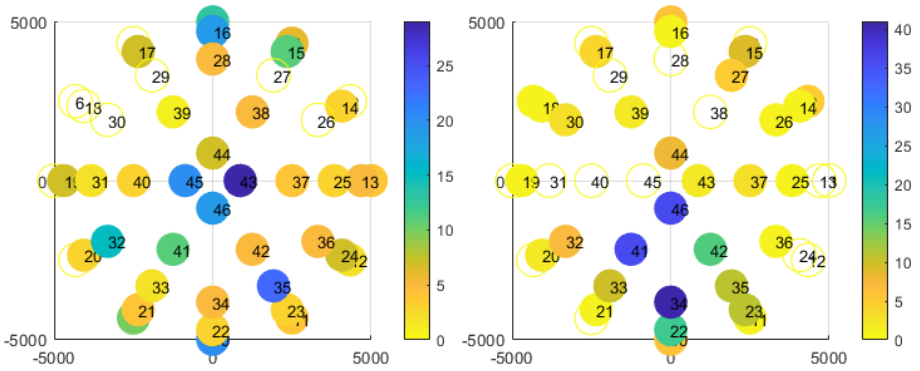
Akcja	Punkty	Pomyłki
Chód osób zdrowych	22 23 26 27 35	18
Chód osób chorych	11 12 14 16 20 22 23 25 27 31 35 37 40 44	45
Stanie	10 14 19 37 39 42	7
Obroty	Wszystkie oprócz 2 5 6 7 8 18 26 27 29 30	295
Schylanie się	2 4 5 6 7 12 17 19 20 22 35 36 37 38 40 41 42	51
Uderzenie	Wszystkie	1447
Kopnięcie niskie	Wszystkie	2303
Kopnięcie w. proste	Wszystkie	8316

Tabela 19: Lista punktów, dla których dochodziło do pomyłek akcji kopnięcie wysokie proste (na 32798 klasyfikacji), wśród zawodników Taekwondo

Akcja	Punkty	Pomyłki
Chód osób zdrowych	3 4 12 15 17 19 24 28 29 32 34 35 37	38
Chód osób chorych	1 3 10 11 12 14 15 16 17 19 22 23 24 25 29 30 31 33 34 35 36 37 38 39 40 42 44 45	138
Stanie	11 23 24 30 33 35 36 37 41 45 46	26
Obroty	Wszystkie oprócz 1 3 5 7 8 9 12 13 24 28 29.. 31 38 40 45	250
Schylanie się	24 34 35 40 41 42 43 46	18
Uderzenie	1 6 7 10 11 14 17 20 22 24 25 34 35 36 37 40 41 42 43 44	70
Kopnięcie niskie	4 5 6 10 11 12 15 16 18 22 23 24 25 26 31 32 34 35 37 38 41 46	105
Kopnięcie w. proste	Wszystkie	3417

Pomyłki z akcjami chód osób zdrowych, stanie, oraz schylanie się ponownie były sporadyczne. Ponownie dotyczyło to 1-2 kopnięć wykonanych przez 1-2 osoby w danym punkcie. W przypadku zawodników Karate równie rzadko dochodziło do pomyłek z akcją chód osób chorych. Kopnięcia zawodników Taekwondo natomiast nieco częściej, ale w dalszym ciągu rzadko były z tą akcją mylone. Błędy te dotyczyły 11 różnych osób - każda z tych osób była mylona w innym punkcie, gdy uderzenie wykonywane było w tarczę.

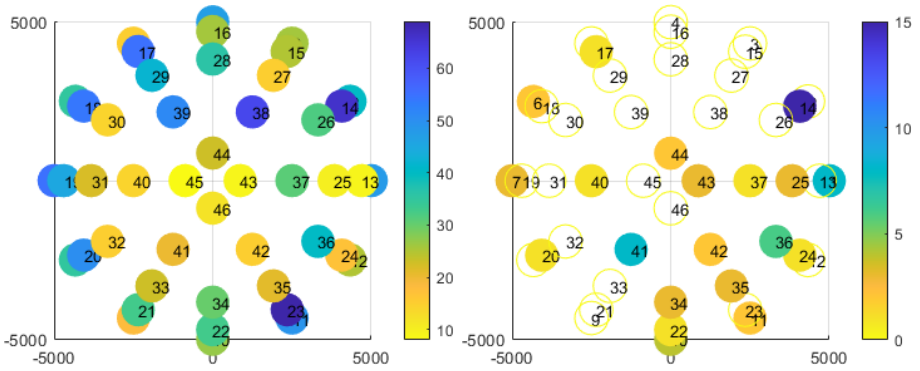
Do pomyłek z akcją obroty dochodziło w większej liczbie punktów. Na rysunku 100 przedstawiono kopułę punktów, w których doszło do pomyłek, gdzie kolor punktu zależy od liczby błędnych klasyfikacji. Kopnięcia zawodników Karate znacznie częściej mylone były z obrotami, gdy wirtualna kamera znajdowała się na szczycie kopuły. W przypadku zawodników Taekwondo do błędów dochodziło częściej, gdy kamera była ustawiona przed zawodnikiem.



(a) (b)

Rysunek 100: Kopuła liczby błędnych klasyfikacji wraz z ich numerami dla sieci LSTM, w których akcja kopnięcie wysokie boczne mylona była z akcją obroty dla zawodników: (a) Karate, (b) Taekwondo.

Do pomyłek z akcją uderzenie, wśród zawodników Karate dochodziło dla każdego położenia wirtualnej kamery. Błędnie klasyfikowane były kopnięcia praktycznie każdego zawodnika widziane, co najmniej z jednej perspektywy. Średnio nagrania danego zawodnika klasyfikowane były błędnie w 15 punktach. Znacznie częściej do pomyłek dochodziło, gdy wirtualna kamera była bliżej podłoża, z prawej strony zawodnika lub za nim (rys. 101 a). Wśród zawodników Taekwondo do pomyłek dochodziło sporadycznie, głównie, gdy wirtualna kamera znajdowała się z lewej strony zawodnika lub przed nim (rys. 101 b).

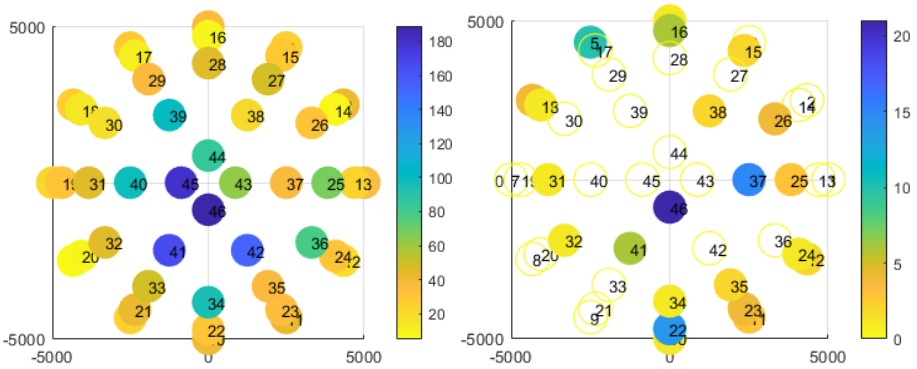


(a) (b)

Rysunek 101: Kopuła liczby błędnych klasyfikacji wraz z ich numerami dla sieci LSTM, w których akcja kopnięcie wysokie boczne mylona była z akcją uderzenie dla zawodników: (a) Karate, (b) Taekwondo.

Podobnie w przypadku pomyłek z akcją kopnięcie niskie - dotyczyły one głównie zawodników Karate. Do pomyłek dochodziło w każdym położeniu wir-

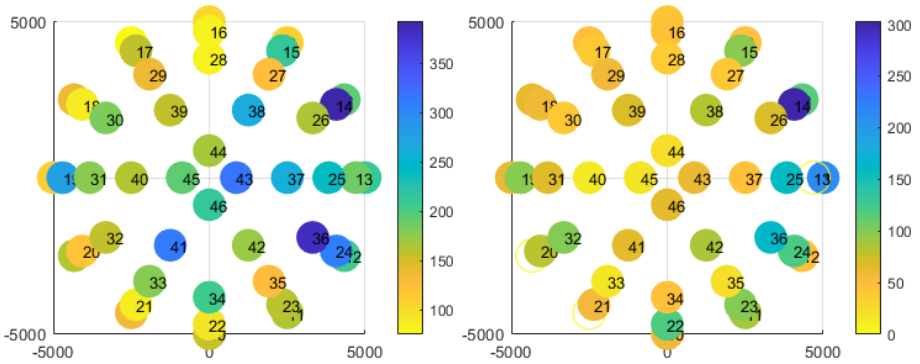
tualnej kamery znacznie częściej, gdy kamera ta ustawiona była blisko szczytu kopuły (rys. 102 a). Ponownie błędy dotyczyły, co najmniej jednego nagrania każdego z zawodników, średnio dla 16 punktów na zawodnika. Może to mieć związek z tym, iż tylko zawodnicy Karate wykonywali akcje kopnięcie niskie - oba rodzaje tych kopnięć są podobne - różnią się tylko wysokością kopnięcia, która widziana z góry jest znacznie cięższa do określenia. Zawodnicy Taekwondo myleni byli dla różnych położeń wirtualnej kamery (rys. 102 b).



(a) (b)

Rysunek 102: Kopuła liczby błędnych klasyfikacji wraz z ich numerami dla sieci LSTM, w których akcja kopnięcie wysokie boczne mylona była z akcją kopnięcie niskie dla zawodników: (a) Karate, (b) Taekwondo.

Akcją, z którą najczęściej mylone było kopnięcie wysokie boczne była akcja kopnięcie wysokie proste. Co najmniej jedno kopnięcie danego zawodnika Karate mylone było z kopnięciem prostym dla minimum 10 różnych położeń wirtualnej kamery (średnio aż 32). Do błędnych klasyfikacji nieco częściej dochodziło, gdy wirtualna kamera znajdowała się z lewej strony zawodnika (rys. 103 a). Pomyłki wśród zawodników Taekwondo również częściej zdarzały się, gdy kamera znajdowała się z lewej strony zawodnika 103 b). Do pomyłek dochodziło praktycznie dla każdego z zawodników, przy czym dla kilku z nich nie doszło do ani jednej pomyłki. Jeśli już dla danego zawodnika dochodziło do błędnej klasyfikacji to dochodziło do niej średnio w 19 różnych punktach.



(a) (b)

Rysunek 103: Kopuła liczby błędnych klasyfikacji wraz z ich numerami dla sieci LSTM, w których akcja kopnięcie wysokie boczne mylona była z akcją kopnięcie wysokie proste dla zawodników: (a) Karate, (b) Taekwondo.

6.4.11 Wnioski

W zależności od akcji można zaobserwować większy lub mniejszy wpływ rzutowania perspektywicznego, na czułość klasyfikacji. Do pomyłek często dochodziło dla tych samych osób, co w przestrzeni trójwymiarowej, przy czym nie dla każdej perspektywy - ruch widziany z danego położenia wirtualnej kamery klasyfikowany był poprawnie, a z innego błędnie.

Oba warianty akcji chód, poza sporadycznymi przypadkami mylone były pomiędzy sobą. W zależności od położenia wirtualnej kamery częściej dochodziło do pomyłek w jedną lub drugą stronę - punkt, w którym czułość klasyfikacji akcji chód osób zdrowych była wysoka, dla akcji chód osób chorych była znacznie niższa. Dodatkowo, ponownie widać wpływ danego schorzenia na poprawność klasyfikacji. Osoby, których schorzenie w znacznymy stopniu upośledza ruch były częściej poprawnie klasyfikowane.

Akcje statyczne ponownie uzyskały najwyższą, czułość klasyfikacji. W przypadku tych akcji widać bardzo wyraźny wpływ położenia wirtualnej kamery nie tylko, na czułość klasyfikacji, ale również na to, z jakimi innymi akcjami dochodziło do pomyłek. Położenie wirtualnej kamery w dużym stopniu determinuje z jaką inna akcją dchodzi do błędnej klasyfikacji.

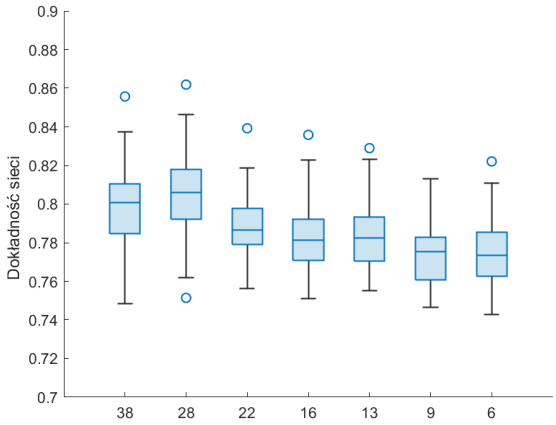
Podobnie w przypadku akcji potencjalnie niebezpiecznych. Ich, czułość klasyfikacji w bardzo dużej mierze zależy od położenia wirtualnej kamery. W skrajnych przypadkach, dla pewnych punktów akcja jest praktycznie zawsze mylona z innymi akcjami czy to statycznymi czy pozostałymi niebezpiecznymi. Można również zauważyć różnicę pomiędzy klasyfikacją zawodników Taekwondo oraz Karate, a także wpływ sposobu wykonania techniki (w powietrze lub w cel), na czułość klasyfikacji.

6.5 Klasyfikacja z wykorzystaniem sieci CNN

6.5.1 Wpływ rozmiaru wektora wejściowego na dokładność sieci

Wpływ rozmiaru wektora wejściowego na ogólną, dokładność klasyfikacji przez sieci CNN danych dwuwymiarowych jest podobny jak w przypadku danych trójwymiarowych. Na rysunku 104 przedstawiono pudełkowy wykres średniej, dokładności sieci we wszystkich 46 położeniach wirtualnej kamery dla różnej liczby markerów w wektorze wejściowym. Można zaobserwować, iż początkowa redukcja znaczników poprawia ogólną, dokładność klasyfikacji, następnie każda kolejna nieznacznie ją obniża. Podobnie jak w przypadku sieci LSTM przy 13 markerach w wektorze wejściowym następuje lekki wzrost średniej.

W stosunku do danych trójwymiarowych poprawie uległa spójność sieci. Średnia różnica pomiędzy najgorszym a najlepszym wynikiem dla danego rozmiaru wektora wejściowego waha się od 11% do 18%.

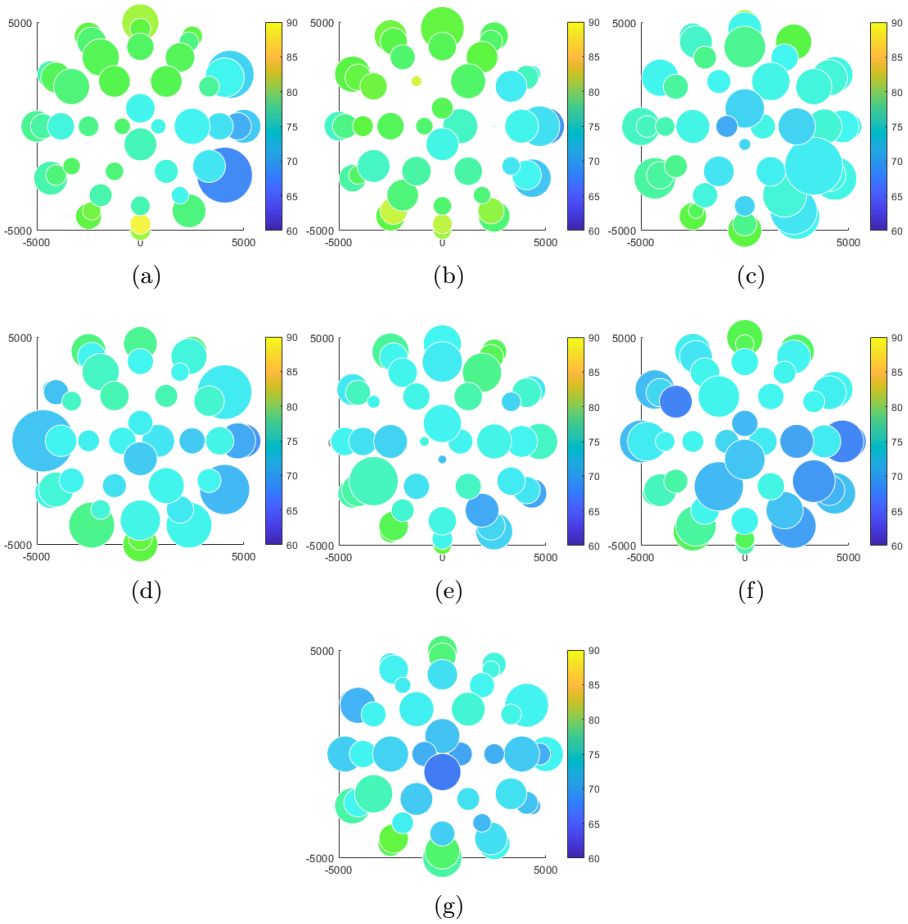


Rysunek 104: Wykres pudełkowy dla średniej dokładności wszystkich sieci LSTM utworzonych dla każdego położenia wirtualnej kamery dla różnej liczby markerów w wektorze wejściowym

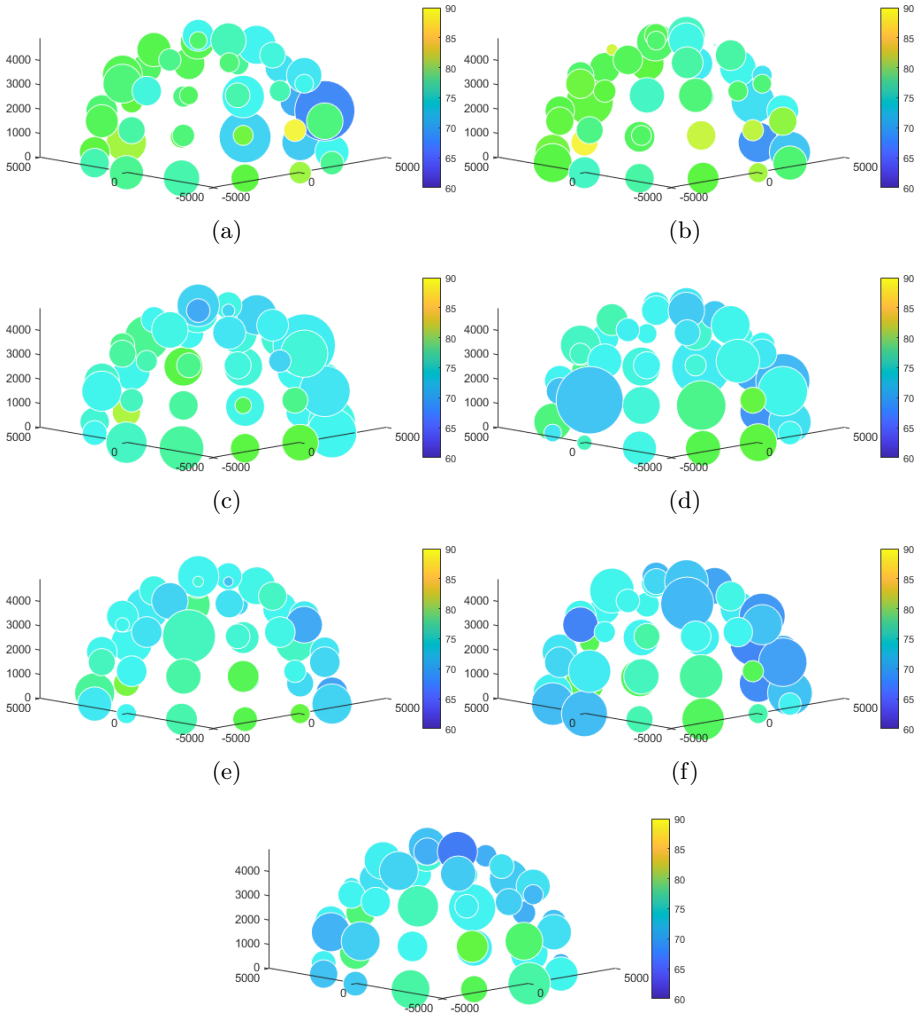
Na rysunkach 105, 106 i 107 przedstawiono kopuły dokładności sieci dla poszczególnych rozmiarów wektora wejściowego. Dodatkowo na rysunku 108 przedstawiono wykresy pudełkowe dokładności, również dla różnych rozmiarów wektora wejściowego z podziałem na lokalizacje wirtualnej kamery (dostępne w pełnej rozdzielczości w dodatku C). Analiza wymienionych wykresów oraz wizualizacji wykazała kilka zależności pomiędzy rozmiarem wektora wejściowego a dokładnością klasyfikacji.

Ponownie zidentyfikowane zostały punkty, w których niezależnie od rozmiaru wielkości wektora wejściowego, dokładność klasyfikacji jest zawsze lepsza/gorsza. Punkty, które zawsze uzyskiwały wyższe wyniki znajdowały się przed lub bezpośrednio za rejestrowaną osobą, stosunkowo blisko podłoża. Do większej liczby błędnych klasyfikacji dochodziło natomiast, gdy wirtualna kamera znajdowała się po lewej stronie osoby (prawa strona kopuły widzianej z góry). Sytuacja ta jest najlepiej widoczna dla wektora wejściowego składającego się ze wszystkich 38 (rys. 105 a, 106 a i 107 a) lub 28 markerów (rys. 105 b, 106 b i 107 b)

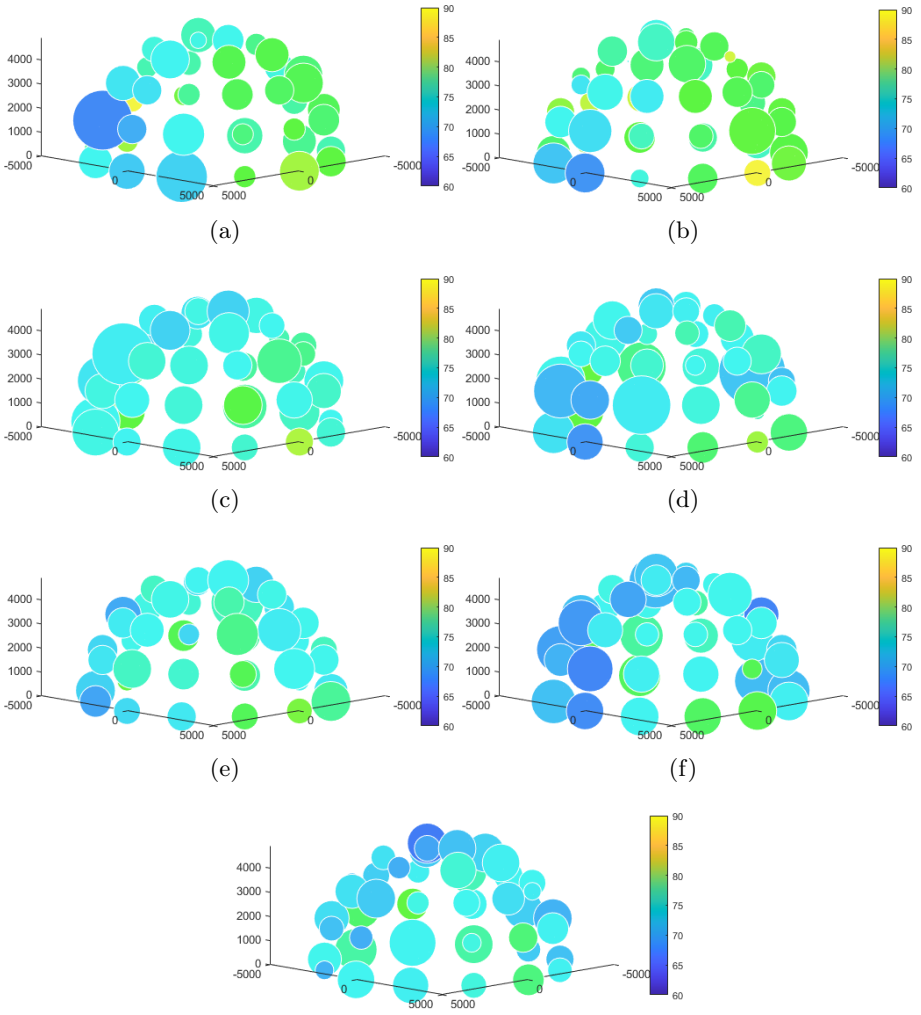
Podobnie jak w przypadku sieci LSTM można zaobserwować wyraźną korelacje pomiędzy spójnością wyników uzyskanych przez sieci, a ich dokładnością w danej lokalizacji wirtualnej kamery. W punktach, które uzyskały średnio lepsze wyniki różnica pomiędzy najgorszym a najlepszym wynikiem zazwyczaj jest mniejsza.



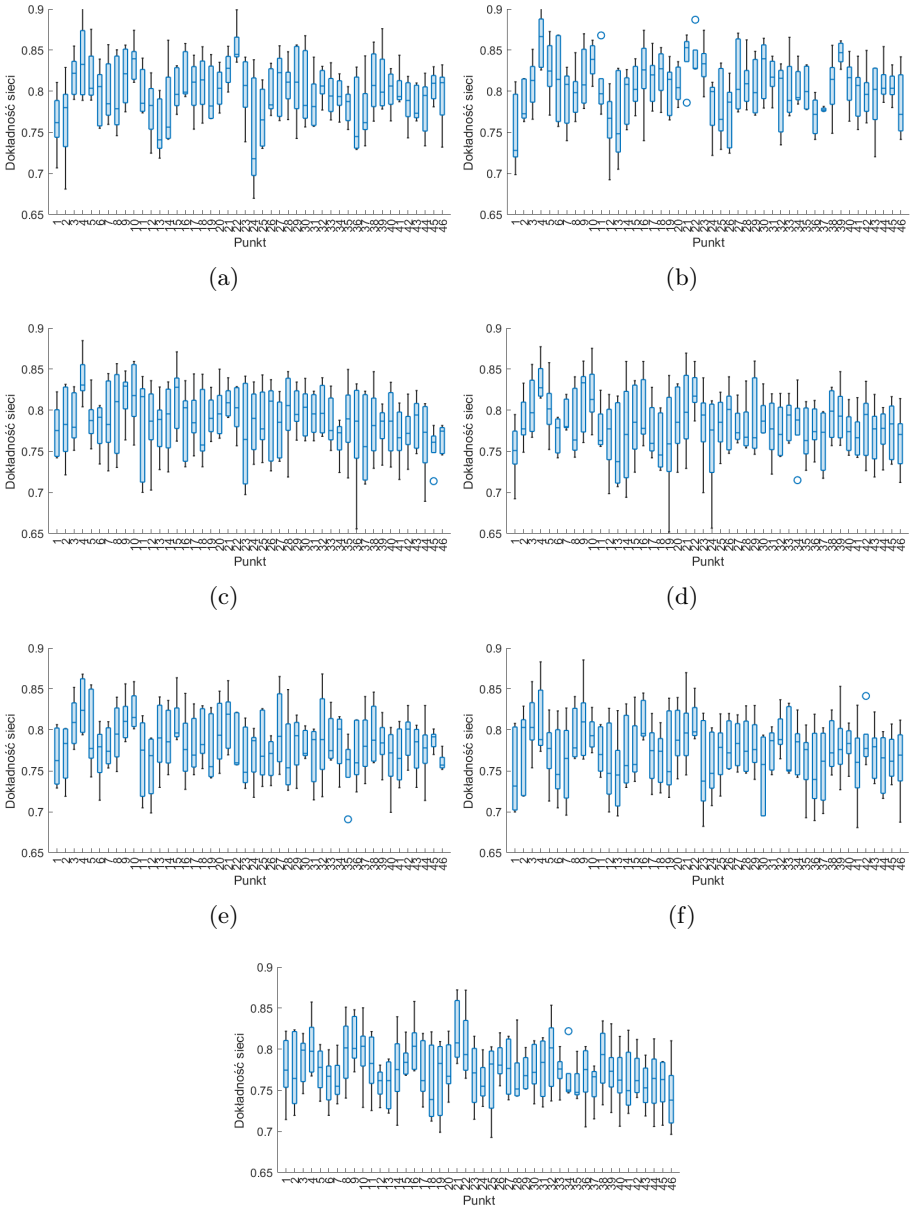
Rysunek 105: Dokładność klasyfikacji sieci CNN dla wszystkich położeń wirtualnej kamery dla różnej liczby znaczników w wektorze wejściowym: (a) 38 znaczników, (b) 28 znaczników, (c) 22 znaczniki, (d) 16 znaczników, (e) 13 znaczników, (f) 9 znaczników, (g) 6 znaczników, widziana z góry.



Rysunek 106: Dokładność klasyfikacji sieci CNN dla wszystkich położeń wirtualnej kamery dla różnej liczby znaczników w wektorze wejściowym: (a) 38 znaczników, (b) 28 znaczników, (c) 22 znaczniki, (d) 16 znaczników, (e) 13 znaczników, (f) 9 znaczników, (g) 6 znaczników, widziana z perspektywy połodniowo-zachodniej.



Rysunek 107: Dokładność klasyfikacji sieci CNN dla wszystkich położeń wirtualnej kamery dla różnej liczby znaczników w wektorze wejściowym: (a) 38 znaczników, (b) 28 znaczników, (c) 22 znaczniki, (d) 16 znaczników, (e) 13 znaczników, (f) 9 znaczników, (g) 6 znaczników, widziana z perspektywy północno-wschodniej.



Rysunek 108: Wykres pudełkowy dokładności sieci CNN dla wszystkich położeń wirtualnej kamery, dla różnej liczby znaczników w wektorze wejściowym: (a) 38 znaczników, (b) 28 znaczników, (c) 22 znaczniki, (d) 16 znaczników, (e) 13 znaczników, (f) 9 znaczników, (g) 6 znaczników.

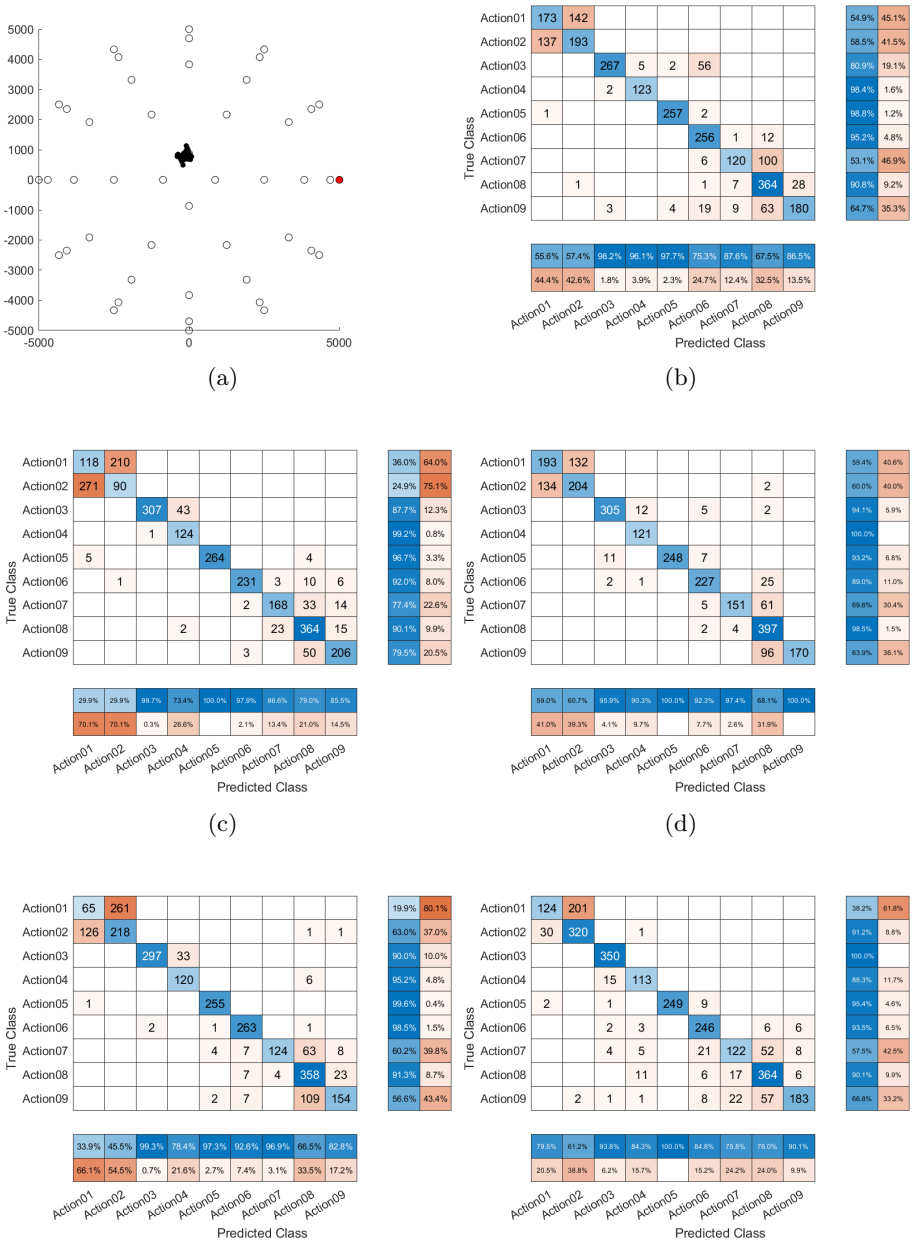
Szczegółowa analiza wszystkich macierzy pomyłek, wykazała, iż zdecydowana większość błędów dotyczyła akcji chód osób zdrowych. Akcja ta mylona była przede wszystkim z akcją chód osób chorych. Dodatkowo dla kilku położeń wirtualnej kamery, dla niektórych walidacji dochodziło do sytuacji gdzie większość osób zdrowych klasyfikowana była, jako chore i odwrotnie - praktycznie wszystkie osoby chore, jako zdrowe. Na rysunku 109 przedstawiono kopułę z zaznaczonym punktem, w którym doszło do omawianej sytuacji wraz z macierzami pomyłek dla wszystkich 5 walidacji dla 13 markerów w wektorze wejściowym. Można zaobserwować, iż wpływ wielkości rozmiaru wektora wejściowego, na czułość klasyfikacji akcji chód jest podobny jak w przypadku ogólnej klasyfikacji. Gorsze rezultaty uzyskano, gdy na wejściu sieci znajdowały się wszystkie markery, następnie z każdą kolejną redukcją wynik ulegał poprawie, a następnie ponownemu pogorszeniu. Najlepsze wyniki zazwyczaj osiągano, gdy w wektorze wejściowym znajdowało się 22, 16 lub 13 markerów.

Podobnie jak w przypadku sieci LSTM, dochodziło również do pomyłek pomiędzy różnymi rodzajami kopnięć, oraz pomiędzy akcjami statycznymi. Zdarzały się również pomyłki pomiędzy grupami - uderzenie klasyfikowane, jako stanie, czy obroty. Do sytuacji tych dochodziło znacznie częściej dla mniejszej liczby markerów w wektorze wejściowym. Na rysunku 110 przedstawiono macierze pomyłek dla tego samego zbioru osób, dla różnej liczby markerów w wektorze wejściowym. W przypadku niektórych położeń wirtualnej kamery możemy dodatkowo zaobserwować poprawę w czułości klasyfikacji jednego z rodzajów kopnięć - kopnięcia niskiego.

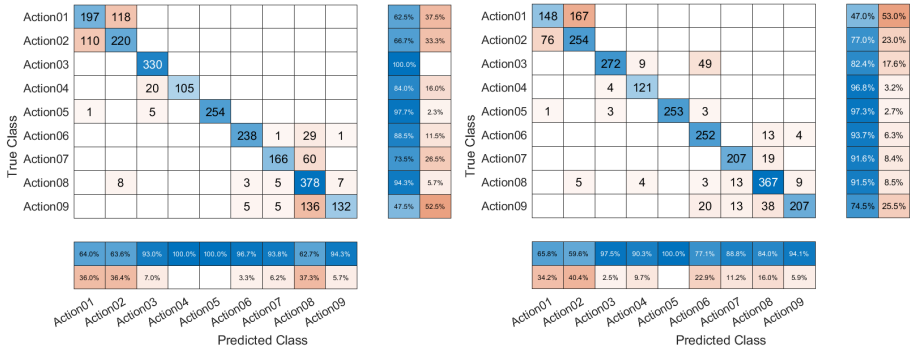
Największy wpływ, na dokładność klasyfikacji miała akcja chód osób zdrowych. Niezależnie od wielkości wektora wejściowego, co najmniej połowa nagrań a maksymalnie aż do 80% było zawsze błędnie klasyfikowanych, jako chód osób chorych. Zmniejszanie liczby markerów na wejściu powodowało, że dla danego położenia wirtualnej kamery procent nagrań błędnie klasyfikowanych rósł.

W przypadku pozostałych akcji wpływ rozmiaru wektora wejściowego, na dokładność klasyfikacji w danym punkcie był znacznie mniejszy. Różnice pomiędzy najlepszymi a najgorszymi wynikami wynoszą kilka procent. Gorsze wyniki uzyskano dla skrajnych rozmiarów wektora wejściowego (38, 9 lub 6 markerów). Najlepsze rezultaty uzyskano dla wektora wejściowego składającego się z 16 znaczników.

Mając powyższe na uwadze, zdecydowano, że dalsza bardziej szczegółowa analiza wpływu położenia wirtualnej kamery, na dokładność klasyfikacji zostanie przeprowadzona dla sieci z 16 markerami na wejściu. W przypadku akcji chód osób zdrowych zostanie dodatkowo przeprowadzona analiza porównawcza wyników uzyskanych, gdy na wejściu sieci było 16 i 28 markerów.

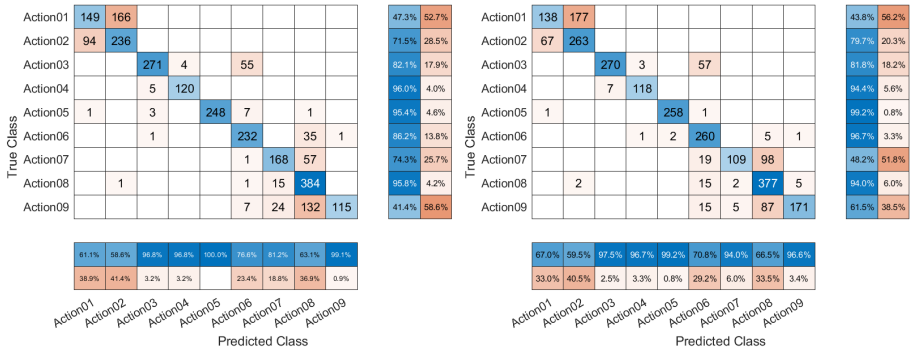


Rysunek 109: Macierze pomyłek dla kolejnych walidacji (b-f) dla wybranego położenia wirtualnej kamery (a), dla 13 markerów w wektorze wejściowym



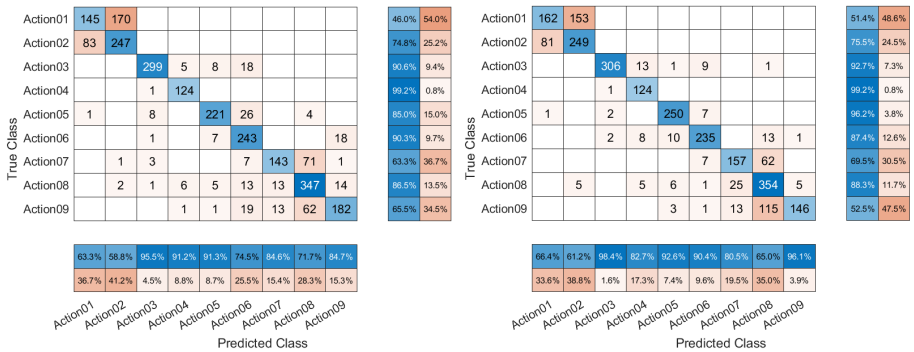
(a)

(b)



(c)

(d)



(e)

(f)

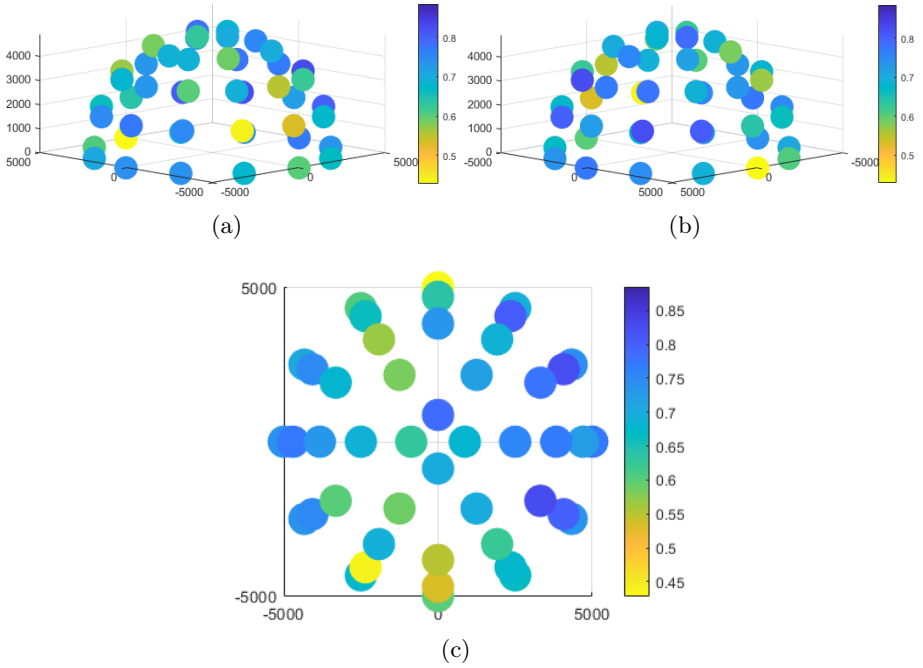
Rysunek 110: Macierze pomyłek dla sieci CNN w wybranym punkcie dla różnej liczby markerów w wektorze wejściowym: (a) 28, (b) 22, (c) 16, (d) 13, (e) 9 i (f) 6.

6.5.2 Akcja Chód osób zdrowych

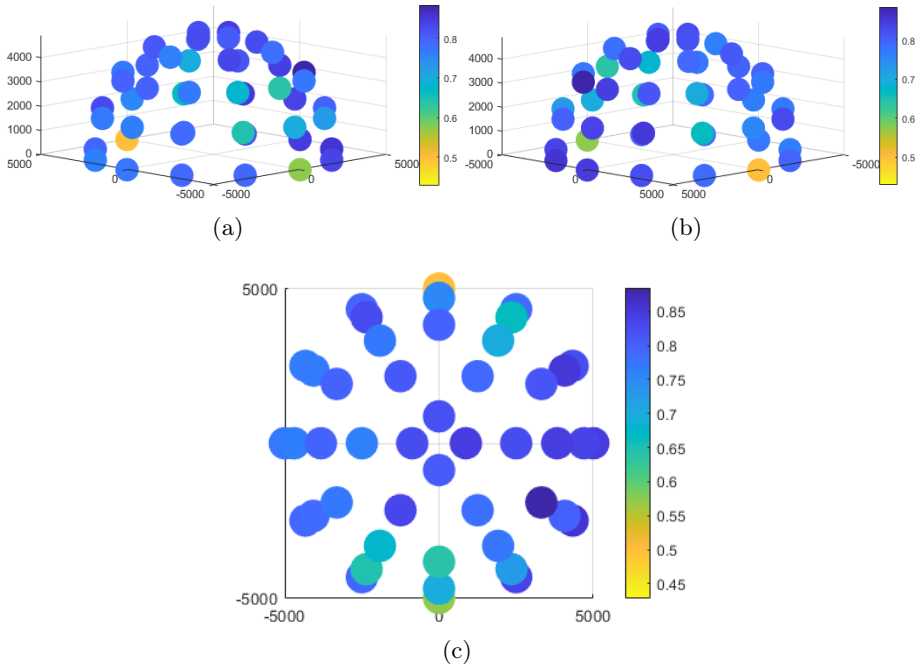
Jak już wcześniej wspomniano, akcja chód osób zdrowych cechowała się najgorszą czułością klasyfikacji. Praktycznie dla każdego położenia wirtualnej kamery, częstotliwość błędów nie spadała poniżej 45%. Akcja ta, okazała się również najbardziej wrażliwa na zmiany wielkości wektora wejściowego. Na rysunkach 111 i 112 przedstawiono kopuły wartości PBA utworzone dla sieci, których wektor wejściowy składał się kolejno z 28 i 16 markerów. W celu lepszej wizualizacji wyników, skale na obu kopułach zostały ujednolicone.

Znacznie lepsze, aczkolwiek w porównaniu do pozostały akcji w dalszym ciągu najgorsze, wyniki uzyskano, gdy w wektorze wejściowym znajdowały się trajektorie 28 znaczników. W zależności od położenia wirtualnej kamery, częstotliwość błędów wynosiła od 42,8% do 82,6%. Przy wektorze wejściowym składającym się z 16 markerów zakres ten wynosił od 49% do 88,5%.

Niezależnie jednak od wielkości wektora wejściowego, punkty, w których czułość klasyfikacji danej akcja była wyższa pozostały takie same. Znajdowały się one na wprost/za rejestrowaną osobą, blisko podłoża. Co warte zaznaczenia, w przypadku sieci LSTM, punkty te również uzyskiwały lepsze wyniki. Dodatkowo dla kilku pozostałych punktów znajdujących się przed nagrywaną osobą również uzyskano nieco lepsze wyniki. W przypadku, gdy w wektorze wejściowym znajdowało się 28 markerów punktów tych było więcej.



Rysunek 111: Kopuła wartości PBA sieci CNN, dla 28 markerów w wektorze wejściowym, dla akcji chód osób zdrowych, widziana z trzech różnych perspektyw (a) z boku, (b) z boku, (c) od góry.



Rysunek 112: Kopuła wartości PBA sieci CNN, dla 16 markerów w wektorze wejściowym, dla akcji chód osób zdrowych, widziana z trzech różnych perspektyw (a) z boku, (b) z boku, (c) od góry.

Niezależnie od wielkości wektora wejściowego dochodziło również do sporadycznych pomyłek z pozostałymi akcjami. W tabelach 20 i 21 przedstawiono listę punktów, w których dochodziło do pomyłek z daną akcją, wraz z łączną liczbą pomyłek w tych punktach kolejno dla wektora wejściowego składającego się z 28 i 16 markerów. Co warte zaznaczenia, jeśli w danym położeniu wirtualnej kamery doszło do błędów niezależnie od wielkości wektora wejściowego, to błędy te dotyczyły najczęściej tych samych osób.

W przypadku pomyłek z akcją chód osób chorych, niezależnie od wielkości wektora wejściowego, co najmniej jedno przejście danej osoby było błędnie klasyfikowane, z co najmniej jednej perspektywy dla 28 markerów w wektorze wejściowym i 4 gdy na wejściu znajdowało się 16 markerów. Liczba punktów, w której dana osoba była, chociaż raz zaklasyfikowana, jako osoba chora była większa, gdy w wektorze wejściowym znajdowało się 16 znaczników.

Podobnie jak w poprzednich eksperymentach, również tym razem osoby, dla których rzadziej dochodziło do błędów nagrywane były w oprogramowaniu Vicon Blade. Przy czym nie znalazła się ani jedna osoba, która w jakimkolwiek punkcie zostałaby w 100% prawidłowo zaklasyfikowana.

Tabela 20: Lista punktów, dla których dochodziło do pomyłek akcji chód osób zdrowych (na 74474 klasyfikacji) z innymi akcjami, przy 28 markerach w wektorze wejściowym

Akcja	Punkty	Pomyłki
Stanie	10 16 22 23 28 33 34	30
Obroty	11 21 35	26
Schylanie się	34	1
Uderzenie	38 39	7
Kopnięcie niskie	32	1
Kopnięcie w. proste	9 15 16 19 21 22 32	15
Kopnięcie w. boczne	24 42	2

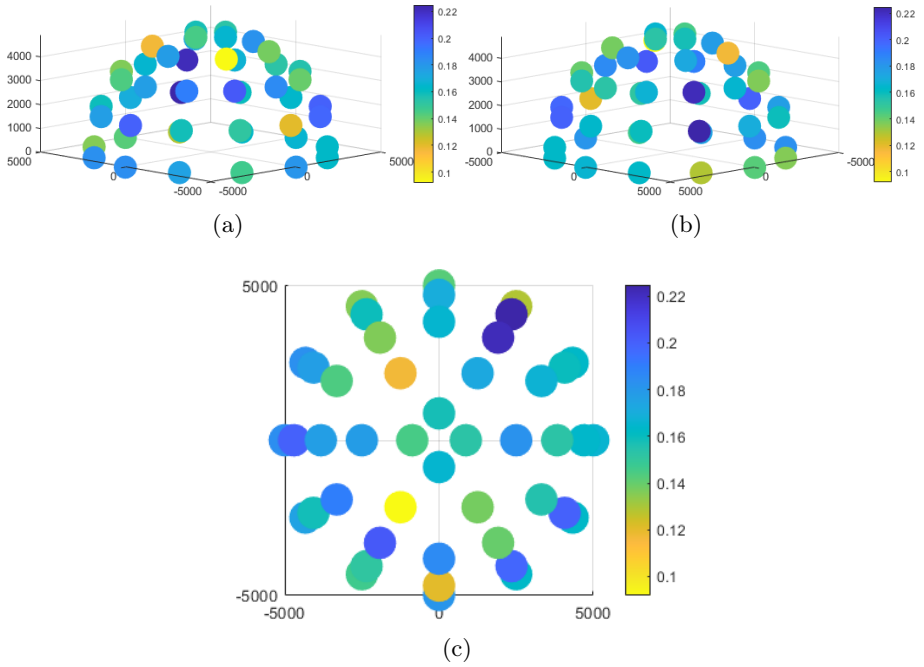
Tabela 21: Lista punktów, dla których dochodziło do pomyłek akcji chód osób zdrowych (na 74474 klasyfikacji) z innymi akcjami, przy 16 markerach w wektorze wejściowym

Akcja	Punkty	Pomyłki
Stanie	10 16 22 28 35	26
Obroty	16 32 33	21
Schylanie się	4 34	2
Uderzenie	15 16	4
Kopnięcie niskie	-	0
Kopnięcie w. proste	3 4 15 16 21 34 35 38 41 42 44 45 46	27
Kopnięcie w. boczne	43	1

6.5.3 Akcja Chód osób chorych

Akcja chód osób chorych cechowała się znacznie większą czułością klasyfikacji niż chód osób zdrowych. Co więcej sieci CNN uzyskały znacznie lepsze rezultaty niż sieci LSTM. W zależności od położenia wirtualnej kamery, czułość klasyfikacji tej akcji wahała się od 77,5% do 90%. Nieco lepsze rezultaty ponownie uzyskiwały punkty znajdujące się przed pacjentem, a gorsze z tyłu po jego lewej stronie lub całkowicie po jego prawej stronie (rys. 113).

W zależności od położenia wirtualnej kamery do błędów dochodziło u różnych osób, w różnym stopniu. Znalazło się 28 osób, ze zwyrodnieniem kręgosłupa, które niezależnie od położenia wirtualnej kamery były zawsze poprawnie klasyfikowane, a 21 spośród nich było też poprawnie klasyfikowane w przestrzeni trójwymiarowej. Było również 7 osób, ze zwyrodnieniem kręgosłupa, które z każdej perspektywy były nieprawidłowo klasyfikowane, a 3 z nich były również zawsze mylone w przestrzeni trójwymiarowej.



Rysunek 113: Kopuła wartości PBA sieci CNN, dla akcji chód osób chorych, widziana z trzech różnych perspektyw (a) z boku, (b) z boku, (c) od góry.

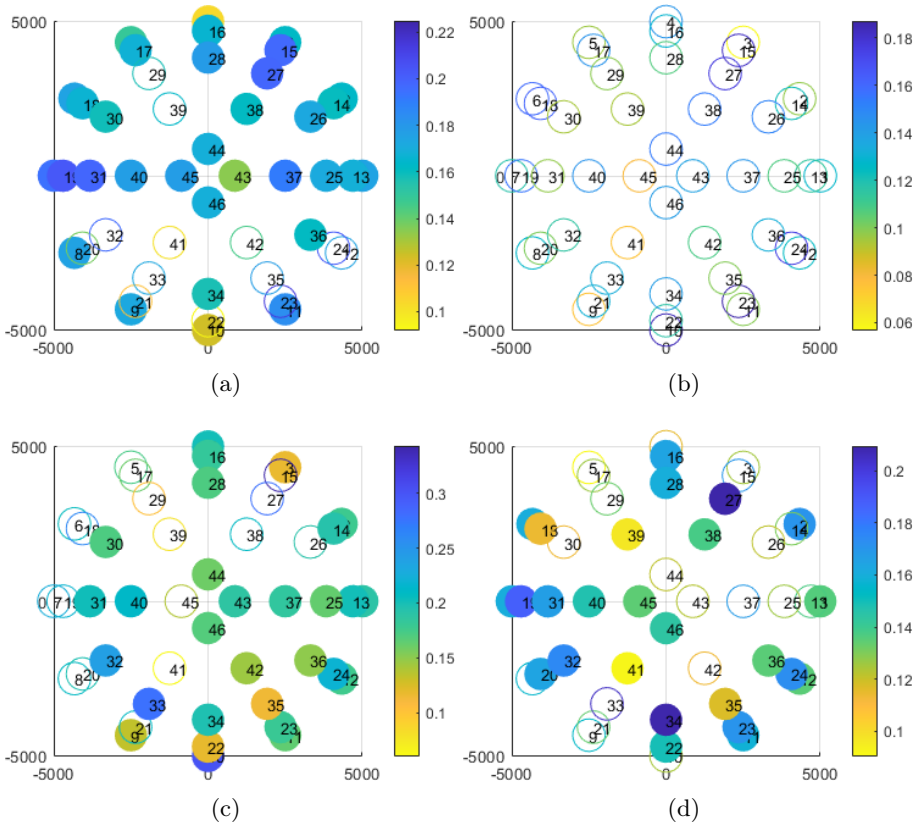
Akcja ta mylona była głównie z akcją chód osób zdrowych. Do pomyłek z akcjami statycznymi dochodziło sporadycznie. Podobnie z akcjami niebezpiecznymi - jeśli już jakieś przejście było mylone to z akcją kopnięcie wysokie proste. Przy czym to, z jaką klasą dochodziło do pomyłek zależało od jednostki chorobowej pacjenta. Na rysunku 114 przedstawiono kopuły PBA, widziane z góry, dla poszczególnych jednostek chorobowych z zaznaczonymi punktami, w których dochodziło do pomyłek z innymi akcjami niż chód osób zdrowych.

Pacjenci po endoprotezoplastyce stawu biodrowego byli myleni z akcją chód osób zdrowych. Wyjątek stanowił jeden pacjent po endoprotezoplastyce obu stawów biodrowych, którego niektóre przejścia widziane z danej perspektywy klasyfikowane były, jako pozostałe akcje. Lista punktów wraz z informacją o klasie, z którą doszło do pomyłki znajduje się w tabeli 22. Co warte zaznaczenia, w przestrzeni trójwymiarowej pacjent ten mylony był tylko z chodem osób zdrowych oraz oboma rodzajami kopnięć wysokich.

Wśród osób, których chód był najczęściej mylony z chodem osób zdrowych znalazło się 17 osób, z czego dla 3 dochodziło do pomyłek w każdej lokalizacji wirtualnej kamery. Osoby te przeszły operacje jednego lub dwóch bioder. Szczegółowa analiza nagrań, w których doszło do pomyłki nie wykazała żadnych widocznych cech wspólnych. Sumarycznie do pomyłek częściej dochodziło, gdy wirtualna kamera znajdowała się po prawej lub lewej stronie pacjenta.

W przypadku osób ze zwyrodnieniem kręgosłupa do pomyłek dochodziło tylko z akcją chód osób zdrowych. Pomyłki dotyczyły 12 pacjentów, z czego chód 7 z nich mylony był, z co najmniej połowy położeń wirtualnej kamery. Do

większej liczby pomyłek dochodziło, gdy kamera znajdowała się z boku pacjenta i była stosunkowo blisko podłoża.



Rysunek 114: Kopuła wartości PBA sieci CNN, dla akcji chód osób chorych, widziana z góry, wraz z numerami punktów, dla różnych jednostek chorobowych (a) endoprotezoplastyka stawu biodrowego, (b) zwyrodnienie kręgosłupa, (c) choroba Parkinsona, (d) udar niedokrwienny mózgu.

Tabela 22: Lista punktów, dla których dochodziło do pomyłek akcji chód osób chorych (na 20654 klasyfikacji) z innymi akcjami, wśród osób po endoprotezoplastyce stawu biodrowego

Akcja	Punkty	Pomyłki
Stanie	16 28	4
Schylanie się	31	2
Uderzenie	8 18 19	6
Kopnięcie niskie	40	1
Kopnięcie w. proste	1 2 3 4 5 6 7 8 9 10 11 13 14 15 16 18 25 26 27 28 30 31 36 37 40 44 45 46	63
Kopnięcie w. boczne	2 11 17 25 26 34 36 37 38 43 45	11

Wśród pacjentów z chorobą Parkinsona do pomyłek z akcjami innymi niż chód osób zdrowych, dochodziło w przypadku dwóch pacjentów. Z czego pomyłki z akcją obroty dotyczyły tylko jednego pacjenta. W tabeli 23 przedstawiono zestawienie akcji, z którymi dochodziło do pomyłek wraz z informacją, w jakim punkcie oraz jak często. Pacjenci Ci uzyskali najwyższy wynik w skali UPDRS, co oznacza, iż objawy choroby były u nich bardzo nasilone. Dodatkowo u pacjentów tych często dochodziło do tzw. zamrożenia chodu, co przypuszczalnie przyczyniło się do pomyłek z pozostałymi klasami. Pomyłki z akcją chód osób zdrowych dotyczyły wszystkich pozostałych pacjentów z tą chorobą. Osoby, które były mylone w ponad połowie punktów uzyskały niższe wyniki w skali UPDRS. Przy czym dla każdego pacjenta znalazła się chodź jedna perspektywa, z której co najmniej jedno jego przejście zostało zaklasyfikowane, jako chód osób zdrowych.

Tabela 23: Lista punktów, dla których dochodziło do pomyłek akcji chód osób chorych (na 12696 klasyfikacji) z innymi akcjami, wśród osób z chorobą Parkinsona

Akcja	Punkty	Pomyłki
Stanie	9 10 16 34	7
Obroty	1 2 3 4 10 11 12 13 14 16 22 23	32
	24 25 28 30 31 32 33 34 37 40 43 46	1
Schylanie się	4	1
Kopnięcie w. proste	10 34 35 36 42 44 46	7

Pacjenci po przebytym udarze niedokrwiennym ponownie myleni byli przede wszystkim z akcją chód osób zdrowych. Wyjątek stanowiły dwie osoby. Dwa przejścia jednego z tych pacjentów widziane z dwóch różnych perspektyw (16 i 34) zaklasyfikowane zostały, jako stanie. Pojedyncze przejścia drugiego pacjenta, w zależności od położenia wirtualnej kamery klasyfikowane były, jako pozostałe akcje, ich lista został przedstawiona w tabeli 24. Do pomyłek z akcją chód osób zdrowych nieco częściej dochodziło, gdy wirtualna kamera znajdowała się z prawej strony pacjenta i była dość blisko ziemi. W grupie tej znalazło się też 8 pacjentów, którzy niezależnie od położenia wirtualnej kamery zawsze byli poprawnie klasyfikowani. Pacjenci Ci chorowali również na dodatkowe schorzenia takie jak problemy z kręgosłupem, cukrzyca czy osteoporoza.

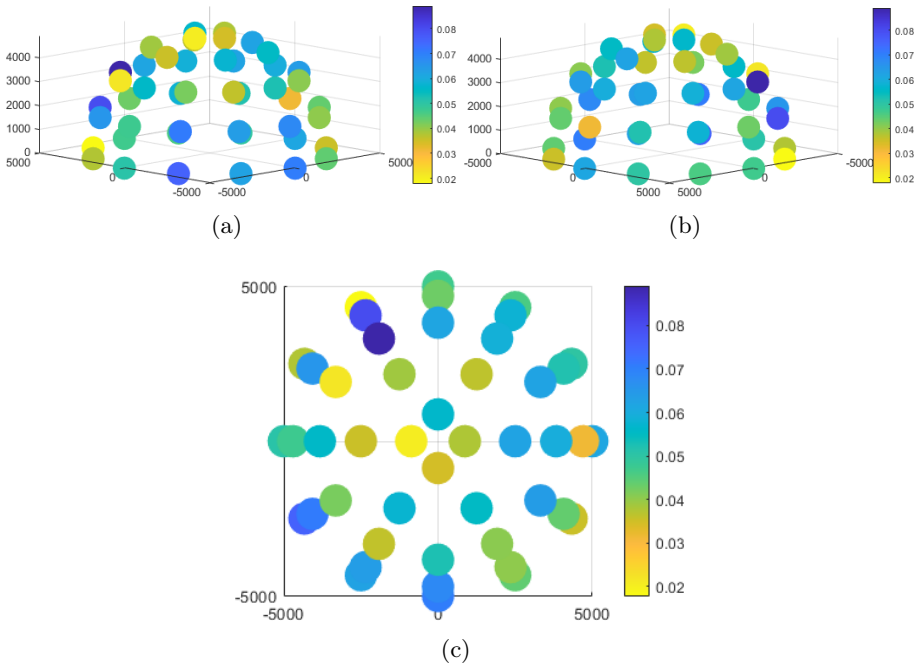
Tabela 24: Lista punktów, dla których dochodziło do pomyłek akcji chód osób chorych (na 13018 klasyfikacji) z innymi akcjami, wśród osób po udarze

Akcja	Punkty	Pomyłki
Stanie	16 22 28 31 32 34 38 45	21
Obroty	28 32	2
Schylanie się	38 39 45	3
Uderzenie	19 20 46	3
Kopnięcie niskie	7	1
Kopnięcie w. proste	1 2 6 11 12 18 20 23 24 27 31 32 34 35 36 40 41 46	31

6.5.4 Akcja Stanie

Akcja stanie charakteryzowała się najwyższą rozpoznawalnością. W zależności od położenia wirtualnej kamery, czułość klasyfikacji tej akcji wahała się od 91% do 98%. Podobnie jak w przypadku sieci LSTM, nieznacznie gorsze rezultaty uzyskały punkty znajdujące się bliżej podłoża (rys. 115).

Akcja ta, w zależności od położenia wirtualnej kamery mylona była przynajmniej z dwoma innymi akcjami. W tabeli 25 przedstawiono listę punktów wraz z informacją o liczbie pomyłek z daną akcją.

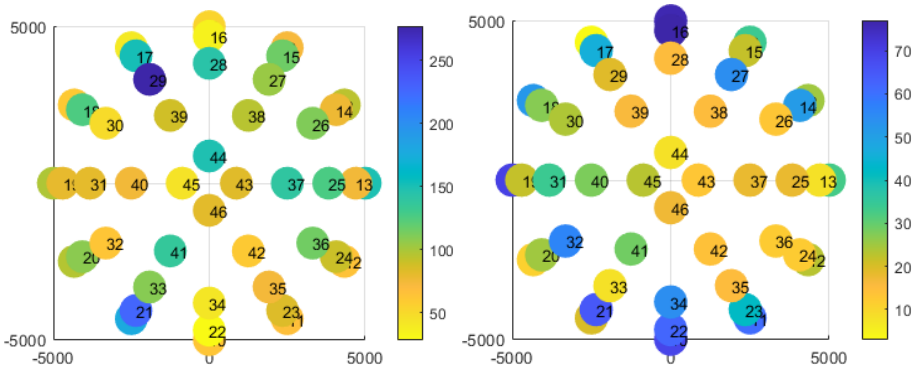


Rysunek 115: Kopuła wartości PBA sieci CNN, dla akcji stanie, widziana z trzech różnych perspektyw (a) z boku, (b) z boku, (c) od góry.

Tabela 25: Lista punktów, dla których dochodziło do pomyłek akcji stanie (na 77464 klasyfikacji) z innymi akcjami

Akcja	Punkty	Pomyłki
Chód osób zdrowych	16 21 22	4
Chód osób chorych	3 4 5 9 10 11 14 15 16 17 22 23 27 30 33 34 35 38 41 43	121
Obroty	Wszystkie	4468
Schylanie się	2 9 13 14 17 22 25 26 27 28 29 30 32 35 37 38 39 40 41 43 44 45	185
Uderzenie	Wszystkie	1439
Kopnięcie niskie	4 8 30 39	7
Kopnięcie w. proste	Wszystkie oprócz 1 4 5 10 11 13 15 16 22 ... 24 25 28 29 30 39 45	257
Kopnięcie w. boczne	34 37 41 45	8

Akcjami, z którymi najczęściej mylona była akcja, stanie były akcje obroty i uderzenie. Pomyłki z obrotami dotyczyły ponad połowy nagrywanych osób, przy czym dana osoba mylona była najczęściej z nie więcej niż 8 różnych położeń wirtualnej kamery. Znalazła się też osoba, której nagrania mylone były w każdym z punktów. Analiza nagrań tej osoby nie wykazała żadnych widocznych odstępstw - brak odrywania stóp od podłoża, rozglądania się czy wymachów rąk. Sumaryczna liczba błędów w poszczególnych punktach została przedstawiona na rysunku 116 a. Do większej liczby błędów dochodziło, gdy wirtualna kamera znajdowała się z lewej strony nagrywanej osoby.



(a) (b)

Rysunek 116: Kopuła liczby błędnych klasyfikacji wraz z ich numerami dla sieci CNN, w których akcja stanie mylona była z akcją (a) obroty, (b) uderzenie.

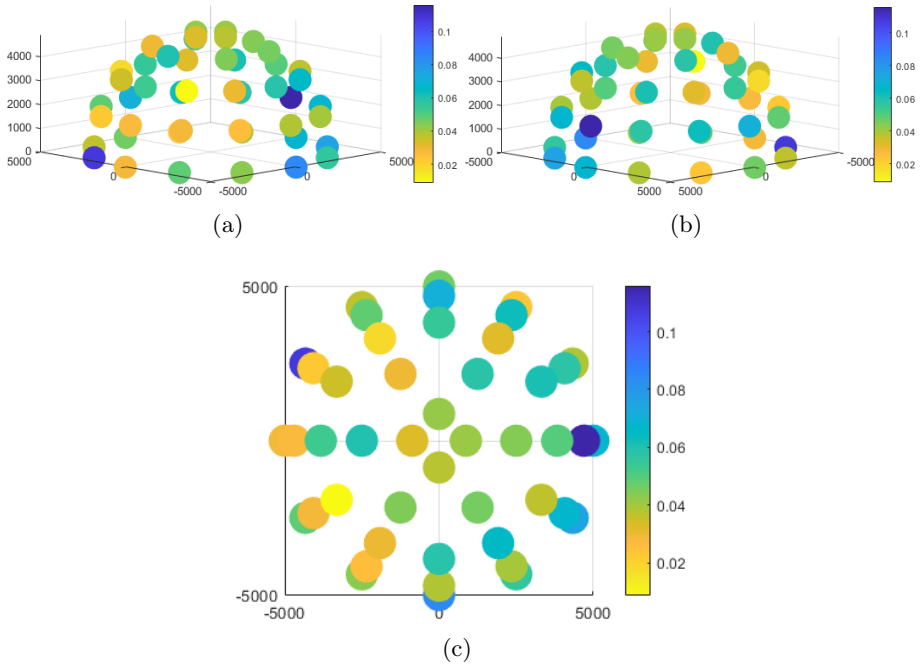
Pomyłki z akcją uderzenie dotyczyły 6 osób. Przy czym 2 osoby mylone były dla każdego położenia wirtualnej kamery, a pozostałe 4 w 3/4 pozycjach. Tym razem w nagraniach osób, u których zawsze dochodziło do pomyłek, zdarzały

się bardziej gwałtowne ruchy rąk, co może być przyczyną nieprawidłowej klasyfikacji. W przypadku tej akcji do błędów częściej dochodziło, gdy wirtualna kamera była przed, za lub z prawej strony rejestrowanej osoby (rys. 116 b).

Co warte zaznaczenia, w zależności od położenia wirtualnej kamery, poszczególne nagrania wspomnianych dwóch osób, oraz jeszcze trzech innych mylone były ze wszystkimi pozostałymi akcjami. Do pomyłek z chodem dochodziło, gdy wirtualna kamera znajdowała się przed lub za nagrywaną osobą, a pomyłki z akcją schylanie w punktach za pacjentem. Do błędów z poszczególnymi rodzajami kopnięć natomiast dochodziło w większości punktów poza tymi znajdującymi się za aktorem lub z jego lewej strony.

6.5.5 Akcja Obroty

Kolejna statyczna akcja - obroty, również charakteryzowała się wysoką, czułością klasyfikacji. Podobnie jak w przypadku danych trójwymiarowych, sieci CNN uzyskały lepsze wyniki. W zależności od położenia wirtualnej kamery, czułość klasyfikacji tej akcji wahała się od 88,5% do aż 99%. Podobnie jak w przypadku sieci LSTM nieco lepsze rezultaty uzyskano, gdy wirtualna kamera znajdowała się po prawej stronie osoby (rys. 117).



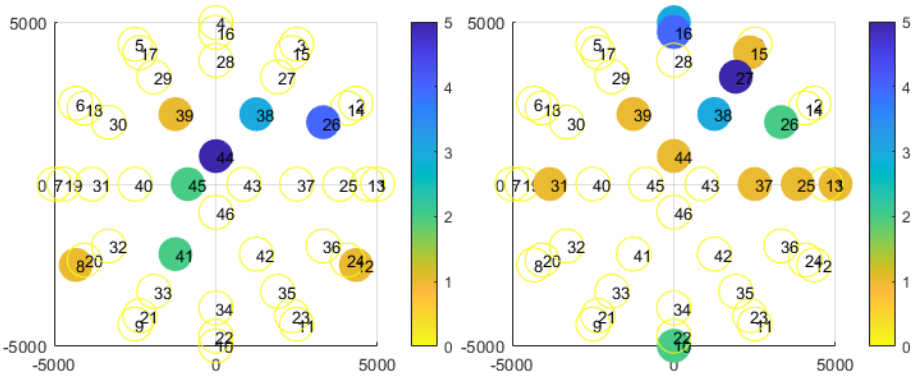
Rysunek 117: Kopuła wartości PBA sieci CNN, dla akcji obroty, widziana z trzech różnych perspektyw (a) z boku, (b) z boku, (c) od góry.

Zdecydowana większość błędów dotyczyła nieprawidłowej klasyfikacji tej akcji, jako stanie. W tabeli 26 przedstawiono zestawienie błędów w poszczególnych lokalizacjach wirtualnej kamery.

Pomyłki z akcją chód osób zdrowych dotyczą tylko jednej osoby. Stanie tej samej osoby było również mylone z chodem osób zdrowych, jednakże w przypadku tych błędów były też dwie inne osoby, które z jednej perspektywy zostały z tą akcją pomylone. Co ważne dochodziło do nich głównie, gdy wirtualna kamera znajdowała się za nagrywaną osobą lub z jej lewej strony (rys. 118). Analiza nagrań wykazała, iż podczas wykonywania tej akcji, osoba, u której głównie dochodziło do pomyłek przemieszczała się najwięcej - obroty nie były w miejscu, a towarzyszyło im przemieszczenie się. Zdecydowana większość osób podczas wykonywania tej akcji nie przemieszczała się.

Tabela 26: Lista punktów, dla których dochodziło do pomyłek akcji obrót (na 28750 klasyfikacji) z innymi akcjami

Akcja	Punkty	Pomyłki
Chód osób zdrowych	8 12 26 38 39 41 44 45	19
Chód osób chorych	1 4 10 15 16 25 26 27 31 37 38 39 44	26
Stanie	Wszystkie oprócz 18 29 45	570
Schylanie się	45	2
Uderzenie	6 14 19 22 34 41 43 45	23
Kopnięcie niskie	34 46	2
Kopnięcie w. proste	1 2 6 8 9 10 12 13 14 15 17 18 19 25 26 31 34 37 38 39 40 41 42 43 44 45	109
Kopnięcie w. boczne	14 25 34 37 43 45	11

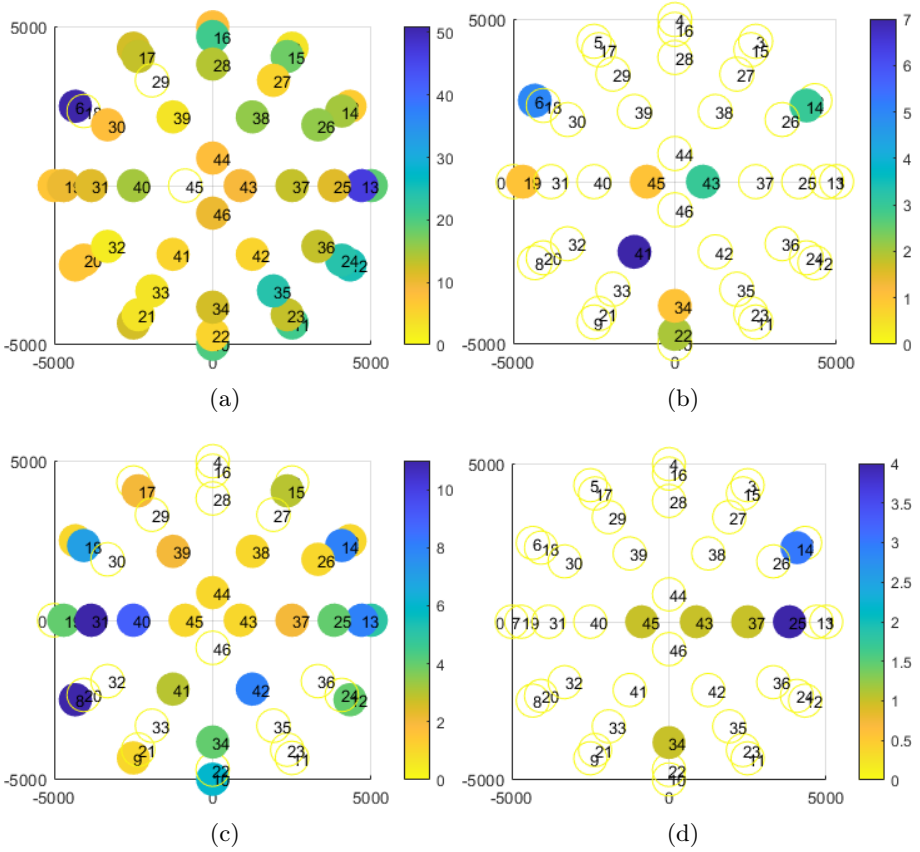


(a) (b)

Rysunek 118: Kopuła liczby błędnych klasyfikacji wraz z ich numerami dla sieci CNN, w których akcja obroty mylona była z akcjami chód osób: (a) zdrowych, (b) chorych.

Błędy z akcją stanie dotyczyły 14 różnych osób, z czego tylko 3 z nich mylone były dla większej liczby położeń wirtualnej kamery. Na rysunku 119 przedstawiono kopułę punktów, których kolor determinuje łączna liczba błędów w danym punkcie z daną akcją. Do pomyłek z akcją stanie dochodziło głównie, gdy wirtualna kamera znajdowała się z lewej strony nagrywanej osoby.

Pozostałe rodzaje błędów ponownie dotyczyły głównie jednej osoby, ze sporadycznymi błędami dla trzech innych (jedno nagranie, każdej z tych osób widziane z 2/3 perspektyw). O ile w przypadku dodatkowego przemieszczania się można zrozumieć pomyłki z akcjami chód osób zdrowych lub chorych, tak wpływ tego ruchu na pomyłki z kopnięciami, zwłaszcza wysokimi, jest niezrozumiały.

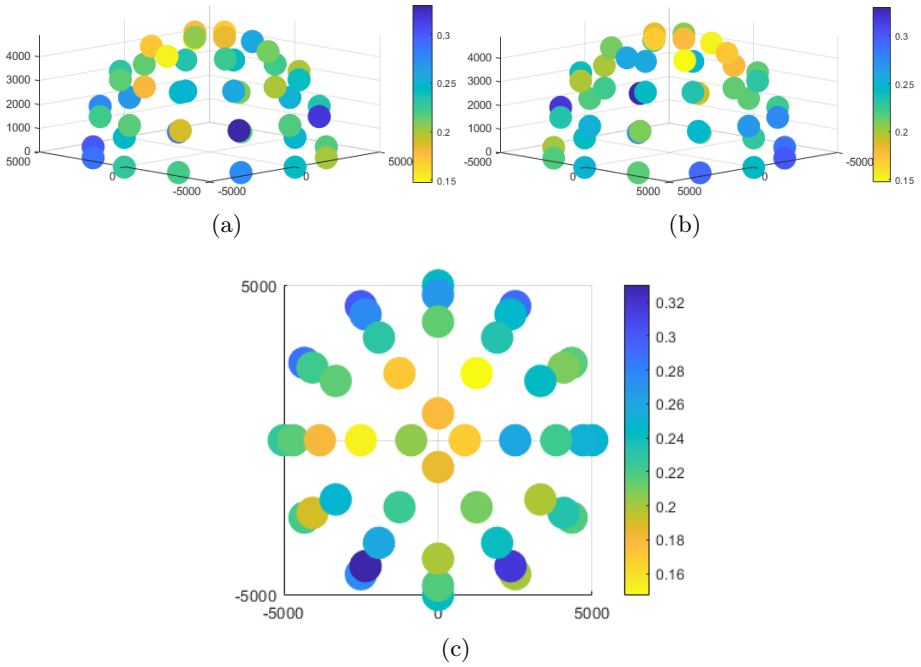


Rysunek 119: Kopuła liczby błędnych klasyfikacji wraz z ich numerami dla sieci CNN, w których akcja obroty mylona była z akcjami: (a) stanie, (b) uderzenie, (c) kopnięcie wysokie proste, (d) kopnięcie wysokie boczne.

6.5.6 Akcja Schylanie się

Podobnie jak w przypadku sieci LSTM, akcja schylanie się charakteryzowała się najgorszą rozpoznawalnością wśród wszystkich akcji statycznych. W zależności od lokalizacji wirtualnej kamery, czułość klasyfikacji wahała się od 67% do 85%. Jednakże w odróżnieniu od sieci LSTM, znacznie lepsze rezultaty uzyskano, gdy wirtualna kamera znajdowała się bliżej szczytu kopuły (rys. 120).

Ponownie dla każdego położenia wirtualnej kamery dochodziło do pomyłek, z co najmniej 4 różnymi akcjami. Akcją, z którą dochodziło do największej liczby pomyłek była akcja stanie, uderzenie i kopnięcie wysokie proste. W tabeli 27 przedstawiono szczegółowe zestawienie pomyłek z poszczególnymi klasami. Dodatkowo na rysunkach 121 i 122 przedstawiono kopuły punktów, w których dochodziło do pomyłek z daną akcją. Kolor danego punktu określa łączna liczba błędów.



Rysunek 120: Kopuła wartości PBA sieci CNN, dla akcji schylanie się, widziana z trzech różnych perspektyw (a) z boku, (b) z boku, (c) od góry.

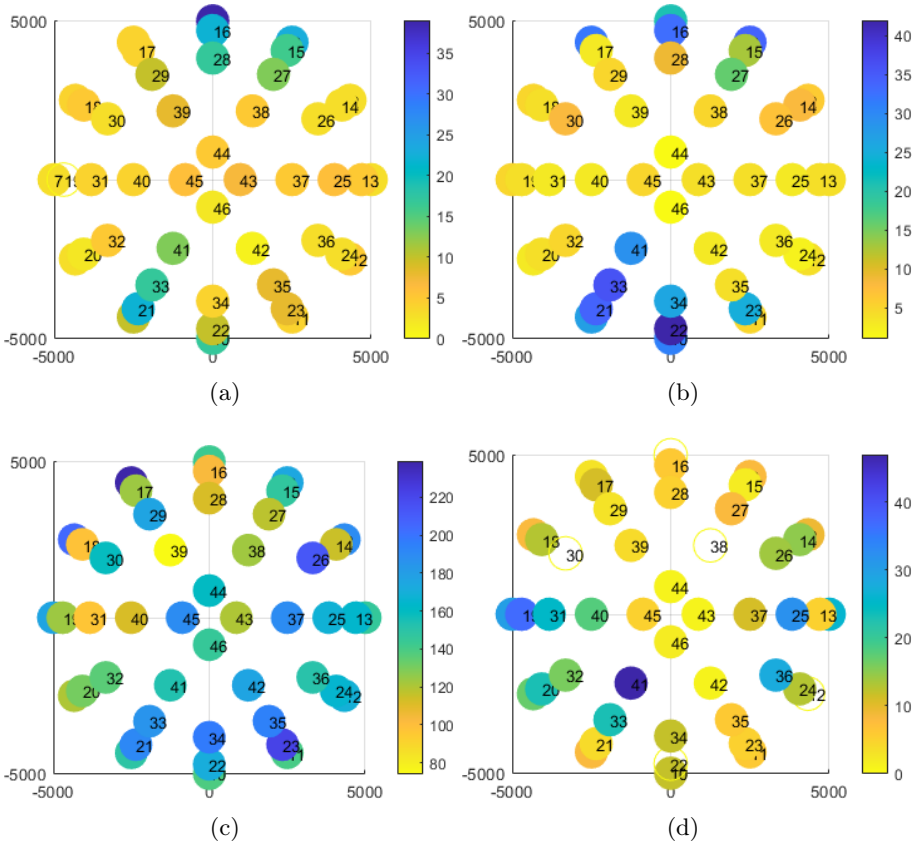
Tabela 27: Lista punktów, dla których dochodziło do pomyłek akcji schylanie się (na 60536 klasyfikacji) z innymi akcjami

Akcja	Punkty	Pomyłki
Chód osób zdrowych	Wszystkie oprócz 19	367
Chód osób chorych	Wszystkie	536
Stanie	Wszystkie	7093
Obroty	Wszystkie - oprócz 4 12 22 30 38	523
Uderzenie	Wszystkie	2461
Kopnięcie niskie	2 5 6 7 9 15 16 17 21 26 27 28 29 31 32 33 36 38 39 43 44	103
Kopnięcie w. proste	Wszystkie	1407
Kopnięcie w. boczne	Wszystkie oprócz 1 3 5 6 11 14 18 24... 2536 37 41 42	163

Do pomyłek z akcją chód osób zdrowych i chód osób chorych dochodziło niezależnie od położenia wirtualnej kamery. Przy czym znacznie częściej, gdy kamera znajdowała się przed lub za osobą i stosunkowo blisko podłoża. Błędy dotyczyły łącznie 115 osób, z czego 23 osoby były mylone zarówno z chodem osób zdrowych jak i chorych. W obu przypadkach, pomyłki dotyczyły tylko 3-4 perspektyw dla danej osoby, przy czym znalazły się osoby, których schylanie się było klasyfikowane, jako chód z ponad połowy perspektyw. Szczegółowa analiza

tych danych w większości przypadków nie wykazała większych nieprawidłowości. U części osób dochodziło do przemieszczania się, jednakże nie było to regułą.

Pomyłki z akcją stanie dotyczyły ponad połowy osób wykonujących tą akcję. Zdecydowana większość z nich mylona była z kilku perspektyw (średnio z 15 różnych). Znalazły się też osoby, których nagrania były mylone z każdej perspektywy. Zazwyczaj skłony te były dość dynamiczne (krótkie), lub powolnie a osoba pozostawała dość długo w skłonie.



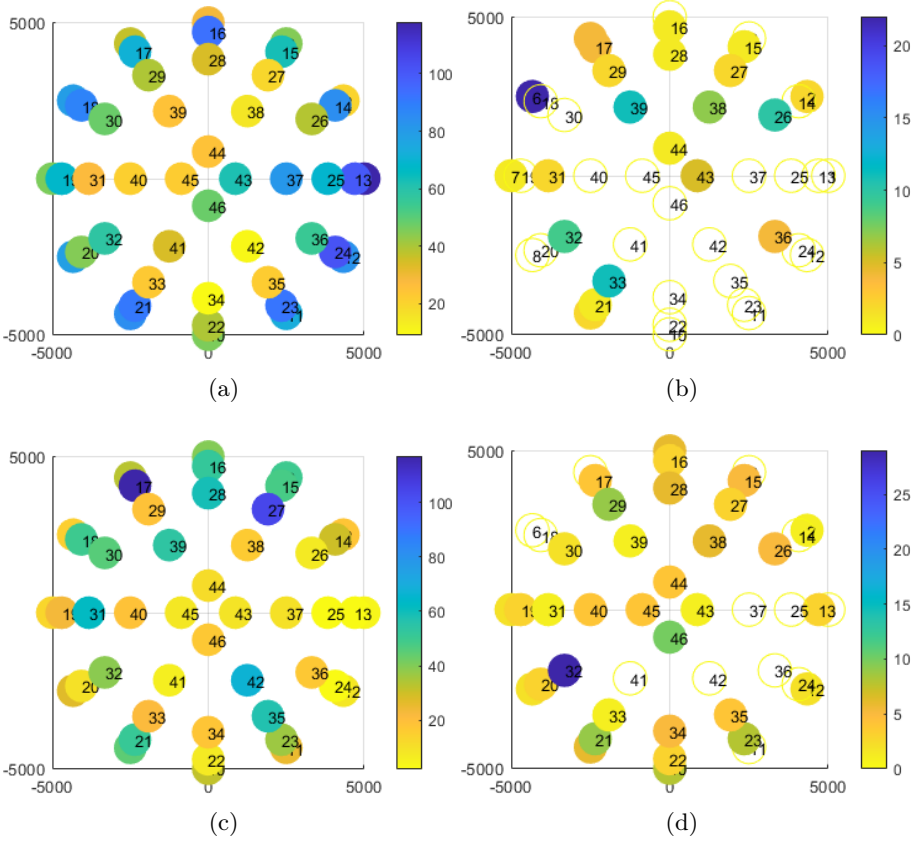
Rysunek 121: Kopuła liczby błędnych klasyfikacji wraz z ich numerami dla sieci CNN, w których akcja schylanie się mylona była z akcjami: (a) chód osób zdrowych, (b) chód osób chorych, (c) stanie, (d) obroty.

Do pomyłek z akcją obroty dochodziło sporadycznie, najczęściej z 3-4 perspektyw dla danej osoby (maksymalnie 20 w przypadku 2 osób). Nieco częściej do pomyłek dochodziło, gdy wirtualna kamera znajdowała się po bokach nagrywanej osoby.

Do błędnych klasyfikacji, jako uderzenie dochodziło we wszystkich położeniach wirtualnej kamery, dla ponad połowy nagrywanych osób. Błędy te znacznie częściej zdarzały się, gdy wirtualna kamera była dość nisko. Średnio każda z osób, u której doszło do pomyłki była błędnie klasyfikowana z 8 różnych perspektyw. Przy czym dla 15 do pomyłek dochodziło w ponad połowie punktów.

Szczegółowa analiza nagrań wykazała, iż pomyłki częściej zdarzały się, gdy skłon wykonywany był bardzo dynamicznie.

Pomyłki z kopnięciem niskim zdarzały się, stosunkowo bardzo rzadko i dotyczyły najczęściej 1-2 perspektyw dla danej osoby. Punkty, w których dochodziło do błędów znajdowały się przede wszystkim za rejestrowaną osobą.

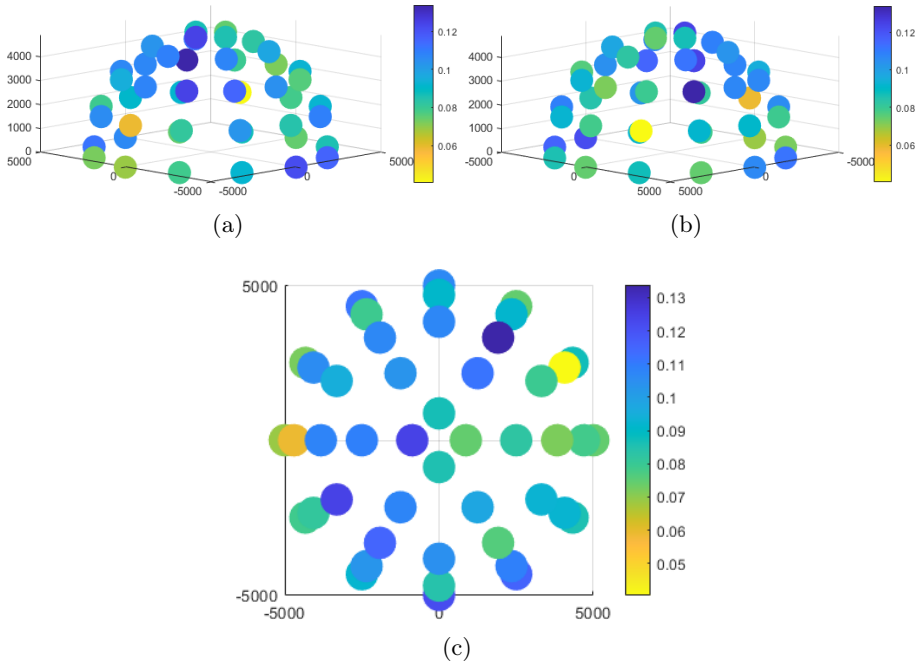


Rysunek 122: Kopuła liczby błędnych klasyfikacji wraz z ich numerami dla sieci CNN, w których akcja schylanie się mylona była z akcjami: (a) uderzenie, (b) kopnięcie niskie, (c) kopnięcie wysokie proste, (d) kopnięcie wysokie boczne.

Podobnie jak w przypadku wcześniejszych akcji, pomyłki z kopnięciami wysokimi dotyczyły przede wszystkim kopnięcia wysokiego prostego. Skłony ponad połowy osób (180) zostały zaklasyfikowane, jako ten rodzaj kopnięcia ze średnio 4 różnych perspektyw (maksymalnie nie więcej jak 20). Do błędów rzadziej dochodziło, gdy wirtualna kamera znajdowała się z lewej strony nagrywanej osoby. Tak jak w przypadku większości omawianych pomyłek dotyczyły one bardziej dynamicznych wykonań tej akcji.

6.5.7 Akcja Uderzenie

Czułość klasyfikacji akcji uderzenie dla większości położeń wirtualnej kamery oscylował wokół 90%. Dla kilku punktów za nagrywanym zawodnikiem, czułość klasyfikacji nieznacznie spadała do 86%, a dla punktów po jego lewej stronie rosła aż do 96% (rys. 123). Do pomyłek dochodziło przede wszystkim z pozostałymi akcjami niebezpiecznymi, w tym głównie z kopnięciem wysokim prostym.



Rysunek 123: Kopuła wartości PBA sieci CNN, dla akcji uderzenie, widziana z trzech różnych perspektyw (a) z boku, (b) z boku, (c) od góry.

Szczegółowa analiza poszczególnych rodzajów błędów w danych położeniach wirtualnej kamery zostanie ponownie przeprowadzona osobno dla zawodników Karate oraz Taekwondo. W tabelach 28 i 29 przedstawiono listę punktów, w których dochodziło do pomyłek z daną akcją, wraz z informacją o sumarycznej liczbie błędów w tych punktach kolejno dla zawodników Karate oraz Taekwondo.

Pomyłki z akcjami chód osób zdrowych oraz chorych zdarzały się bardzo sporadycznie, a w przypadku zawodników Taekwondo do pomyłek z chodem osób zdrowych nie doszło ani razu.

Do błędnej klasyfikacji, jako stanie dochodziło w przypadku zawodników obu dyscyplin. W przypadku zawodników Karate, punkty, w których dochodziło do największej liczby błędów znajdowały się przed zawodnikiem, lub za nim lekko z jego prawej strony (rys. 124 a). Błędy te dotyczyły pojedynczych uderzeń w tarczę 17 zawodników. Każdy z nich był błędnie klasyfikowany średnio z 2-3 perspektyw. W przypadku zawodników Taekwondo do błędów częściej dochodziło, gdy wirtualna kamera znajdowała się przed zawodnikiem (rys. 124 b). Błędy do-

tyczyły, co najmniej jednego uderzenia w deskę, większości zawodników, średnio z 3 perspektyw.

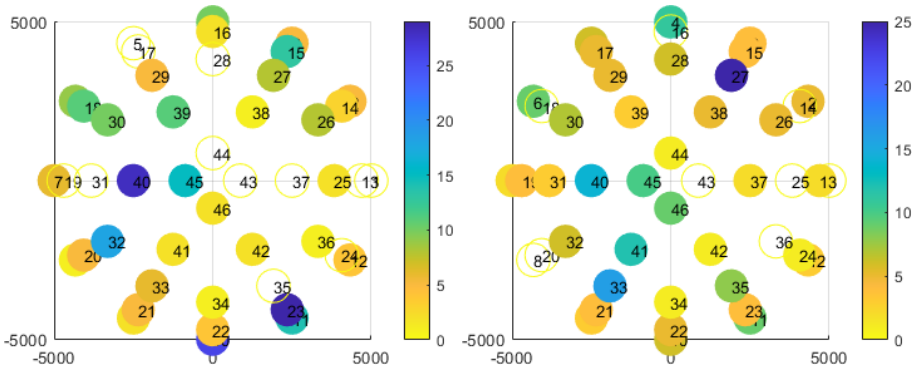
Tabela 28: Lista punktów, dla których dochodziło do pomyłek akcji uderzenie (na 27692 klasyfikacji) z innymi akcjami, wśród zawodników Karate

Akcja	Punkty	Pomyłki
Chód osób zdrowych	43	1
Chód osób chorych	40	2
Stanie	Wszystkie - [1 5 13 17 19 24 28 31 35... 37 43 44]	274
Obroty	Wszystkie - [3 6 9 32 34 39 42 44]	294
Schylanie się	4 5 11 12 13 14 19 22 24 25 31 37 40	70
Kopnięcie niskie	Wszystkie - [1 3 7 12 17 18 20 24 27... 28 42]	110
Kopnięcie w. proste	Wszystkie	2196
Kopnięcie w. boczne	Wszystkie	744

Tabela 29: Lista punktów, dla których dochodziło do pomyłek akcji uderzenie (na 32338 klasyfikacji) z innymi akcjami, wśród zawodników Taekwondo

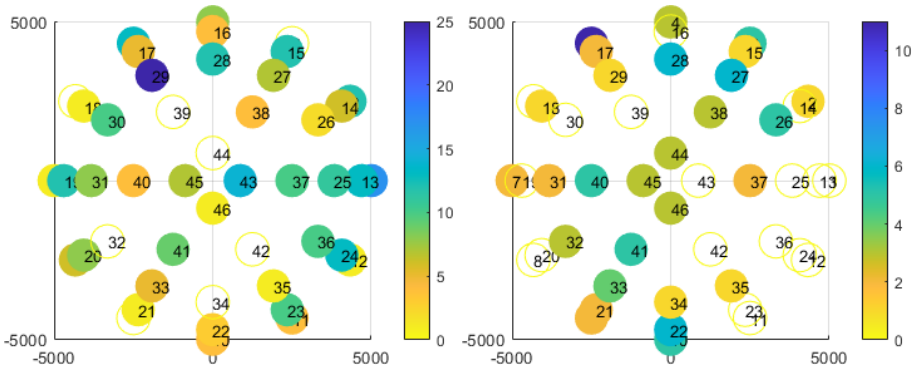
Akcja	Punkty	Pomyłki
Chód osób chorych	10 21 30 34 41 43	8
Stanie	Wszystkie - [1 8 14 16 18 20 25 36 43]	226
Obroty	Wszystkie - [1 6 8 11 12 13 14 16 19 20... 23 24 25 30 36 39 42 43]	94
Schylanie się	5 9 12 13 22 24 25 28 29 33 35 39 40 43	39
Kopnięcie niskie	2 4 13 22 30 31 34 35 41 46	20
Kopnięcie w. proste	Wszystkie - [1 6 11 13 16 17 19 25 26]	308
Kopnięcie w. boczne	2 34 35 36 42 43 45 46	18

Błędy z akcją obroty zdarzały się dużo częściej dla zawodników Karate niż Taekwondo. Na rysunku 125 przedstawiono kopuły punktów, w których dochodziło do błędów. W przypadku zawodników Karate do mniejszej liczby pomyłek dochodziło, gdy wirtualna kamera znajdowała się przed zawodnikiem. Błędy dotyczyły 1-2 uderzeń połowy zawodników widzianych średnio z 4 różnych perspektyw. W nagraniach, które zostały błędnie sklasyfikowane zawodnicy uderzali zarówno w tarczę jak i w powietrze. Wśród zawodników Taekwondo do błędnych klasyfikacji nie dochodziło, gdy kamera znajdowała się przed zawodnikiem lub z jego lewej strony. Ponownie pomyłki dotyczyły jednego bądź dwóch nagrań 15 różnych zawodników widzianych z kilku perspektyw. Średnio uderzenia jednego zawodnika były błędnie klasyfikowane dla 5 różnych położeń wirtualnej kamery.



(a) (b)

Rysunek 124: Kopuła liczby błędnych klasyfikacji wraz z ich numerami dla sieci CNN, w których akcja uderzenie mylona była z akcją stanie dla zawodników (a) Karate, (b) Taekwondo.



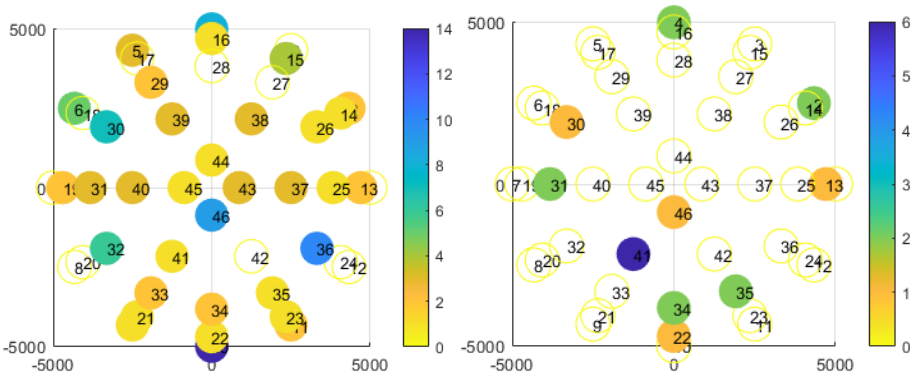
(a) (b)

Rysunek 125: Kopuła liczby błędnych klasyfikacji wraz z ich numerami dla sieci CNN, w których akcja uderzenie mylona była z akcją obroty dla zawodników (a) Karate, (b) Taekwondo.

Do błędnej klasyfikacji uderzeń, jako schylania się dochodziło znacznie rzadziej. W przypadku zawodników Karate głównie, gdy wirtualna kamera znajdowała się po bokach zawodnika, nieco częściej z jego lewej strony. Błędy ponownie dotyczyły pojedynczych uderzeń w tarczę widzianych z 1-2 różnych perspektyw. Podobnie w przypadku zawodników Taekwondo - błędna klasyfikacja zdarzała się dla danej osoby maksymalnie w 3 różnych położeniach wirtualnej kamery i dotyczyła uderzeń w tarczę.

Pomyłki z akcją kopnięcie niskie zdarzały się głównie w przypadku zawodników Karate. Dotyczyły uderzeń zarówno w tarczę jak i w powietrze. Punkt, w którym dochodziło do największej sumarycznej liczby pomyłek znajdował się

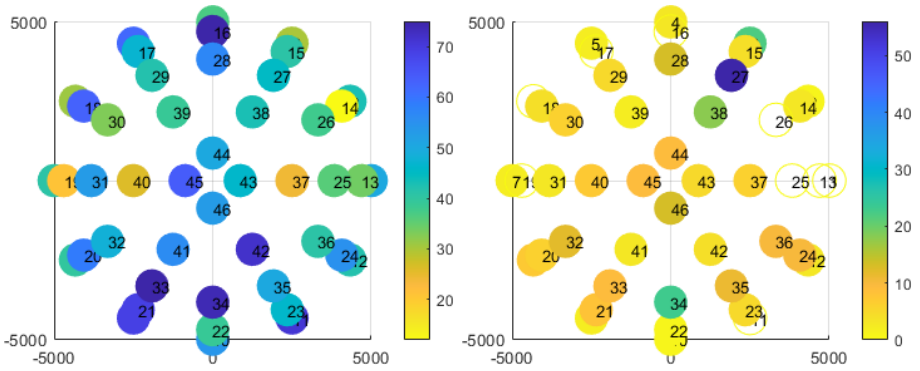
przed zawodnikiem, blisko podłoża. W przypadku zawodników obu dyscyplin błędy dotyczyły kilku osób widzianych średnio z 2/3 różnych perspektyw.



(a) (b)

Rysunek 126: Kopuła liczby błędnych klasyfikacji wraz z ich numerami dla sieci CNN, w których akcja uderzenie mylona była z akcją kopnięcie niskie dla zawodników (a) Karate, (b) Taekwondo.

Akcją, z którą najczęściej mylone było uderzenie była ponownie akcja kopnięcie wysokie proste. Co najmniej jedno uderzenie w tarczę zawodników Karate, z co najmniej jednej, a średnio 15 różnych perspektyw zostało błędnie zaklasyfikowane, jako kopnięcie wysokie proste. Do pomyłek dochodziło w podobnej liczbie z każdej perspektywy, nieznacznie częściej, gdy wirtualna kamera znajdowała się za/przed zawodnikiem (rys. 127 a). W przypadku zawodników Taekwondo pomyłki były znacznie rzadsze. Dochodziło do nich głównie, gdy wirtualna kamera znajdowała się za zawodnikiem lekko z jego prawej strony (rys. 127 b). Błędy te dotyczyły uderzeń w tarczę kilku zawodników. Każdy z nich mylony był najczęściej z 5 różnych perspektyw.

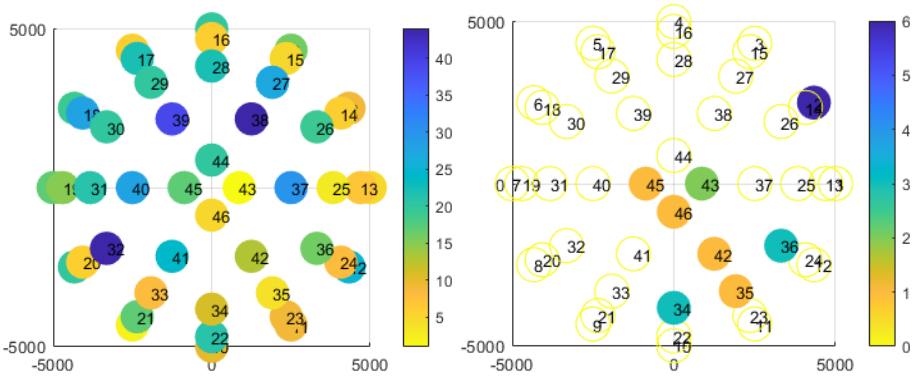


(a) (b)

Rysunek 127: Kopuła liczby błędnych klasyfikacji wraz z ich numerami dla sieci CNN, w których akcja uderzenie mylona była z akcją kopnięcie wysokie proste dla zawodników (a) Karate, (b) Taekwondo.

Do pomyłek z akcją kopnięcie wysokie boczne wśród zawodników Karate znacznie rzadziej dochodziło, gdy wirtualna Kamera znajdowała się z lewej strony zawodnika (rys. 128 a). Błędy ponownie dotyczyły głównie tych samych uderzeń w tarczę. Dane uderzenie danego zawodnika widziane z jednej perspektywy klasyfikowane było, jako kopnięcie wysokie boczne a z innej, jako kopnięcie wysokie proste. Średnia liczba punktów, w której dochodziło do pomyłek dla danego zawodnika wynosiła 13 (minimalnie 1 maksymalnie 24). Błędne klasyfikacje wśród zawodników Taekwondo, ponownie, zdarzały się bardzo rzadko. Dotyczyły tylko czterech zawodników, średnio dla 3 różnych położeń wirtualnej kamery. Punkty, w których dochodziło do błędów znajdowały się przed zawodnikiem lekko z jego lewej strony.

Podobnie jak w przypadku poprzednich eksperymentów, widać bardzo wyraźny wpływ zarówno sposobu wykonania uderzenia (tarcza lub powietrze), oraz tego, jaką dyscyplinę sportową uprawiała dana osoba zarówno na czułość klasyfikacji jak i rodzaj akcji, z którą dochodzi do pomyłek.



(a) (b)

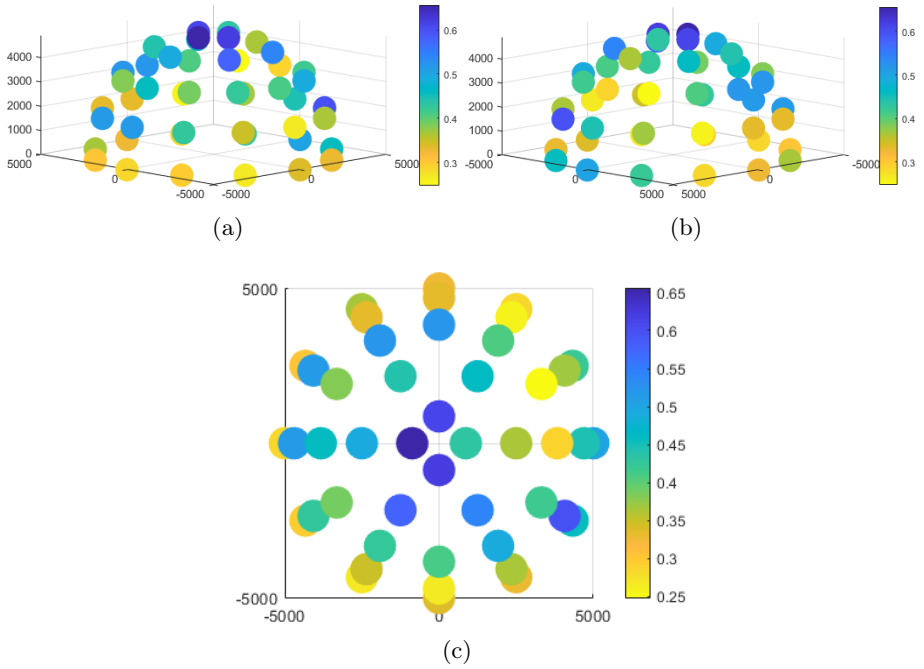
Rysunek 128: Kopuła liczby błędnych klasyfikacji wraz z ich numerami dla sieci CNN, w których akcja uderzenie mylona była z akcją kopnięcie wysokie boczne dla zawodników (a) Karate, (b) Taekwondo.

6.5.8 Akcja Kopnięcie niskie

Czułość klasyfikacji akcji kopnięcie niskie jest drugą najniższą po akcji chód osób zdrowych. Dodatkowo jest ona bardzo zależna od położenia wirtualnej kamery - waha się od 34% do 75% (rys. 129). Do błędów znacznie częściej dochodziło gdy wirtualna kamera zjadowała się wyżej - im bliżej szczytu kopuły tym gorsze rezultaty.

Akcja ta, najczęściej mylona była z pozostałymi akcjami niebezpiecznymi (uderzenie i oba warianty kopnięć wysokich), jednakże zdarzały się też pomyłki z pozostałymi akcjami. W tabeli 30 przedstawiono zestawienie punktów, w których dochodziło do pomyłek z daną akcją wraz z sumaryczną liczbą pomyłek w wymienionych punktach.

Pomyłki z akcjami chód osób zdrowych oraz chorych zdarzały się bardzo sporadycznie. Dotyczyły kopnięcia w tarczę, kilku osób. Co więcej każda z tych osób była mylona w innym punkcie, przy czym wszystkie te punkty znajdowały się po lewej stronie zawodnika lub za nim.

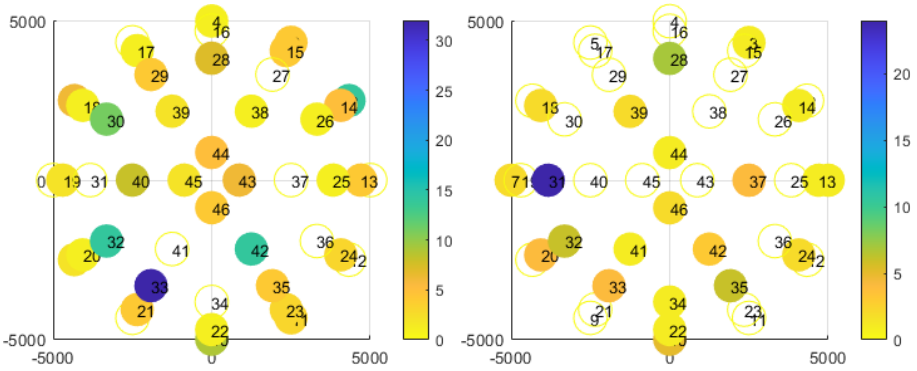


Rysunek 129: Kopuła wartości PBA sieci CNN, dla akcji kopnięcie niskie, widziana z trzech różnych perspektyw (a) z boku, (b) z boku, (c) od góry.

Tabela 30: Lista punktów, dla których dochodziło do pomyłek akcji kopnięcie niskie (na 49818 klasyfikacji) z innymi akcjami

Akcja	Punkty	Pomyłki
Chód osób zdrowych	34 37 43	4
Chód osób chorych	2 6 13 15 16 27 36 37 38 39 42 43	30
Stanie	Wszystkie oprócz 1 5 7 9 12 16 27 31 34... 36 37 41	184
Obroty	Wszystkie oprócz 9 22	273
Schylanie się	1 3 7 10 13 14 18 20 22 24 28 31... 32 33 34 35 37 39 41 42 44 46	81
Uderzenie	Wszystkie	1376
Kopnięcie w. proste	Wszystkie	15336
Kopnięcie w. boczne	Wszystkie	2858

Do błędnej klasyfikacji z akcją stanie dochodziło dla praktycznie każdego zawodnika widzianego średnio z 5 różnych perspektyw. Nieco częściej gdy wirtualna kamera znajdowała się z prawej strony zawodnika (rys. 130 a). Kopnięcia które były błędnie klasyfikowane w większości były kopnięciami w tarczę, przy czym błędy dotyczyły też kopnięć w powietrze.

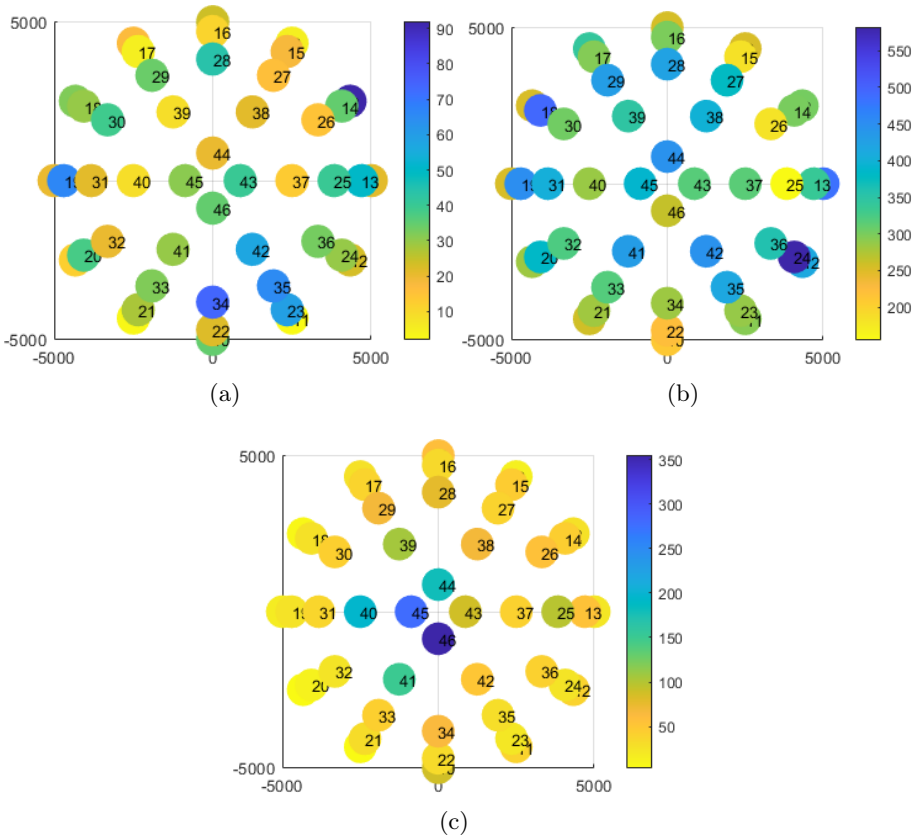


(a) (b)

Rysunek 130: Kopuła liczby błędnych klasyfikacji wraz z ich numerami dla sieci CNN, w których akcja kopnięcie niskie mylona była z akcjami: (a) stanie, (b) schylanie się.

Do pomyłek z akcją schylanie dochodziło rzadziej, głównie gdy wirtualna kamera była stosunkowo blisko podłoża. Błędna klasyfikacja dotyczyła uderzenia w powietrze widzianego z 2-3 różnych perspektyw.

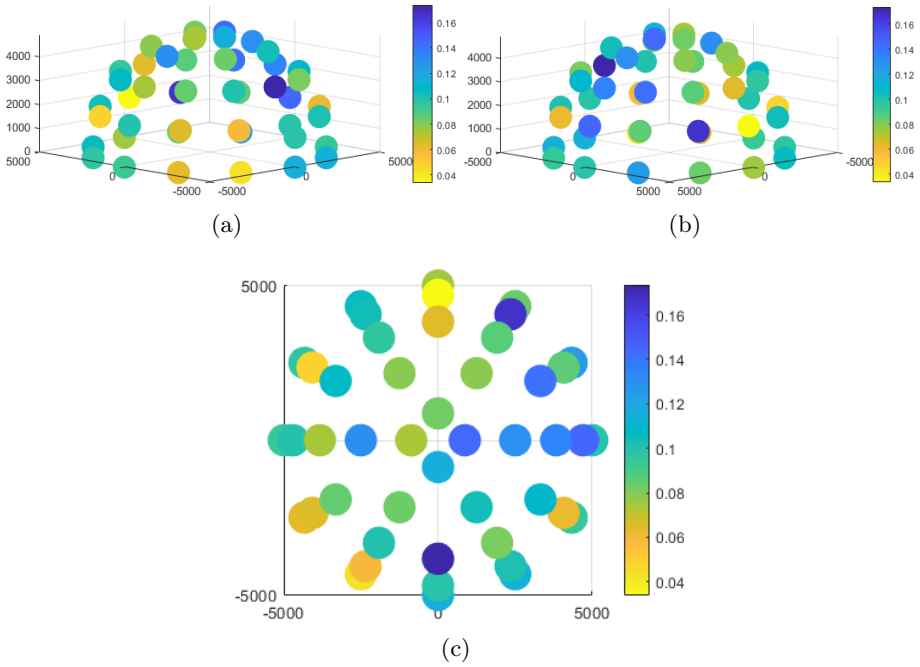
Kopnięcie niskie każdego z zawodników Karate było błędnie klasyfikowane jako uderzenie średnio dla 14 perspektyw, jako kopnięcie wysokie proste z 35, a wysokie boczne z 16. Do pomyłek z akcją uderzenie rzadziej dochodziło gdy wirtualna kamera znajdowała się za zawodnikiem (rys. 131 a). Na nagraniach, w których dochodziło do błędów zawodnicy w większości kopali w tarczę. Do błędnej klasyfikacji jako kopnięcie wysokie boczne najczęściej dochodziło gdy wirtualna kamera znajdowała się tuż nad zawodnikiem (rys. 131 b). Akcją z którą najczęściej mylone było kopnięcie niskie ponownie była akcja kopnięcie wysokie proste. Do błędów dochodziło z każdej perspektywy, nieco rzadziej gdy wirtualna kamera znjdowała się dość nisko (rys. 131 c). Praktycznie wszystkie nagrania, które zostały błędnie sklasyfikowane zawierały kopnięcie w powietrze.



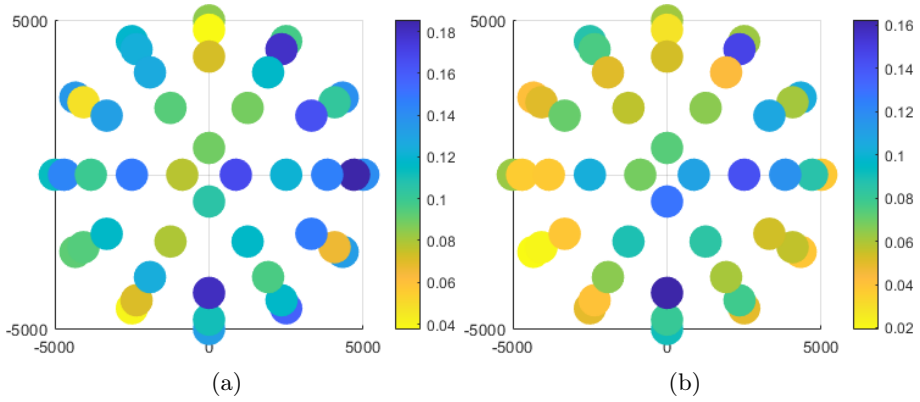
Rysunek 131: Kopuła liczby błędnych klasyfikacji wraz z ich numerami dla sieci CNN, w których akcja kopnięcie niskie mylona była z akcjami: (a) uderzenie, (b) kopnięcie wysokie proste, (c) kopnięcie wysokie boczne.

6.5.9 Akcja kopnięcie wysokie proste

Czułość klasyfikacji kopnięcia wysokiego prostego była dość dobra i wahała się od 82% do 96,5%. Nieco lepsze rezultaty uzyskiwano, gdy wirtualna kamera znajdowała się z prawej strony zawodnika (rys. 132). Do błędów wśród zawodników Karate dochodziło nieco częściej, gdy wirtualna kamera znajdowała się z lewej strony zawodnika (rys. 133 a). W przypadku zawodników Taekwondo do większej sumarycznej liczby pomyłek dochodziło, gdy wirtualna kamera znajdowała się przed zawodnikiem lekko z jego lewej strony (rys. 133 b). Szczegółowe listy punktów, w których dochodziło do błędów z daną klasą dla danego sportu walki zostały przedstawione w tabelach 31 i 32.



Rysunek 132: Kopuła wartości PBA sieci CNN, dla akcji kopnięcie wysokie proste, widziana z trzech różnych perspektyw (a) z boku, (b) z boku, (c) od góry.



Rysunek 133: Kopuła wartości PBA sieci CNN, dla akcji kopnięcie wysokie proste, widziana z góry dla zawodników (a) Karate, (b) Taekwondo.

W przypadku zawodników Karate do pomyłek z akcją chód osób zdrowych nie doszło ani razu. Błędne klasyfikowanie, jako chód osób chorych zdarzało się bardzo sporadycznie, głównie, gdy wirtualna kamera znajdowała się za zawodnikiem lekko z jego prawej strony. Pomyłki te dotyczyły czterech różnych

zawodników - kopnięcia każdego z nich mylone były z innej perspektywy. W przypadku zawodników Taekwondo dochodziło do błędnych klasyfikacji zarówno z chodem osób zdrowych jak i chorych. W obu przypadkach pomyłki te były sporadyczne, średnio dany zawodnik był mylony z jedną z tych akcji tylko z 1-2 perspektyw. Większość zawodników mylona była z chodem osób chorych. W nagraniach, które były błędnie klasyfikowane zawodnicy najczęściej dodatkowo przemieszczali się po wykonaniu kopnięcia.

Podobnie w przypadku błędnej klasyfikacji z akcją stanie - znacznie częściej dochodziło do niej w przypadku zawodników Taekwondo. Czterej zawodnicy Karate myleni byli tylko wtedy, gdy kamera znajdowała się przed zawodnikiem. Zawodnicy Taekwondo myleni byli z prawie połowy perspektyw, przy czym ponownie kopnięcie danego zawodnika było błędnie klasyfikowane w nie więcej jak 3 punktach. Kopnięcia, które mylone były ze staniem były wykonywane głównie w tarczę.

W przypadku akcji schylanie się, do pomyłek dochodziło bardzo rzadko. Zazwyczaj błędy dotyczyły jednego zawodnika widzianego z jednej perspektywy.

Tabela 31: Lista punktów, dla których dochodziło do pomyłek akcji kopnięcie wysokie proste (na 29532 klasyfikacji)

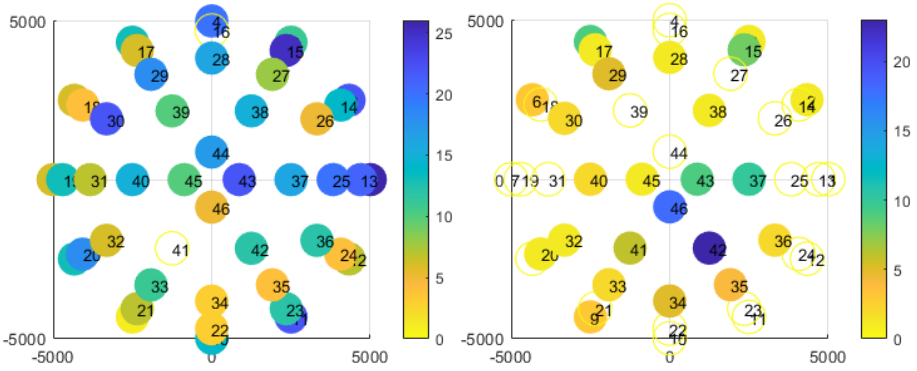
Akcja	Punkty	Pomyłki
Chód osób chorych	17 23 28 29 46	22
Stanie	22 33 34 42 46	17
Obroty	Wszystkie oprócz 16 41	542
Schylanie się	4 22 23 25 29	9
Uderzenie	Wszystkie	911
Kopnięcie niskie	Wszystkie	912
Kopnięcie w. boczne	Wszystkie	853

Tabela 32: Lista punktów, dla których dochodziło do pomyłek akcji kopnięcie wysokie proste (na 62652 klasyfikacji), wśród zawodników Taekwondo

Akcja	Punkty	Pomyłki
Chód osób zdrowych	15 26	2
Chód osób chorych	2 4 6 7 8 12 13 14 15 21 22 25 26 28 29 32 33 35 36 37 38 39 41 42 43 44 45	91
Stanie	2 5 6 12 13 14 15 22 26 27 30 32 33 34 35 40 41 42 43 44 46	72
Obroty	2 3 5 6 9 15 17 20 28 29 30 32 33 34 35 36 37 38 40 41 42 43 45 46	119
Schylanie się	3 5 12 13 15 17 22 25 33 34 43	26
Uderzenie	2 3 4 5 6 8 12 13 14 15 17 19 20 22 23 24 25 26 30 32 34 35 36 37 40 41 42 43 46	259
Kopnięcie niskie	Wszystkie oprócz 1 8 9 16 20 28 29 31	277
Kopnięcie w. boczne	Wszystkie	1947

Pomyłki z akcją obroty zdarzały się znacznie częściej, w porównaniu do

pozostałych akcji statycznych. W przypadku zawodników Karate do pomyłek nieco rzadziej dochodziło, gdy wirtualna kamera znajdowała się przed zawodnikiem (rys. 134 a). Błędy dotyczyły 14 różnych zawodników, kopnięcia każdego z nich mylone były w średnio 13 różnych punktach. W przypadku zawodników Taekwondo średnia liczba punktów, w których dochodziło do pomyłek na zawodnika wynosiła 5. Punkty, w których dochodziło do największej liczby błędów znajdowały się bliżej szczytu kopuły (rys. 134 b). W przypadku zawodników obu dyscyplin, błędnie klasyfikowane były kopnięcia w powietrze.



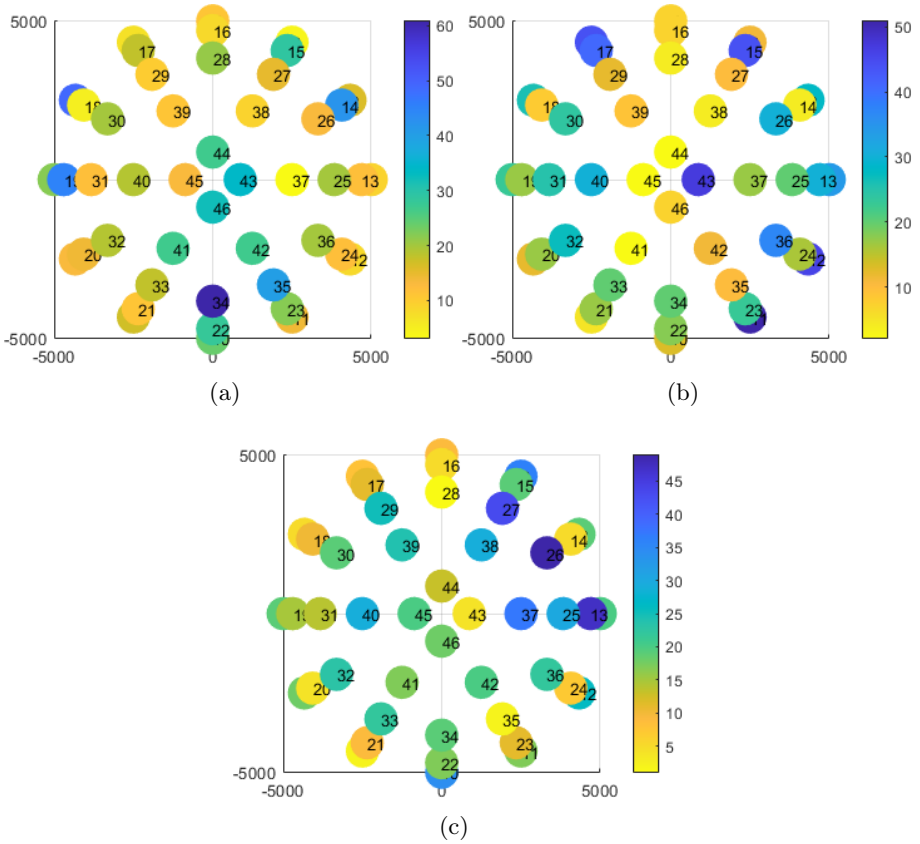
(a) (b)

Rysunek 134: Kopuła liczby błędnych klasyfikacji wraz z ich numerami dla sieci CNN, w których akcja kopnięcie wysokie proste mylona była z akcją obroty dla zawodników: (a) Karate, (b) Taekwondo.

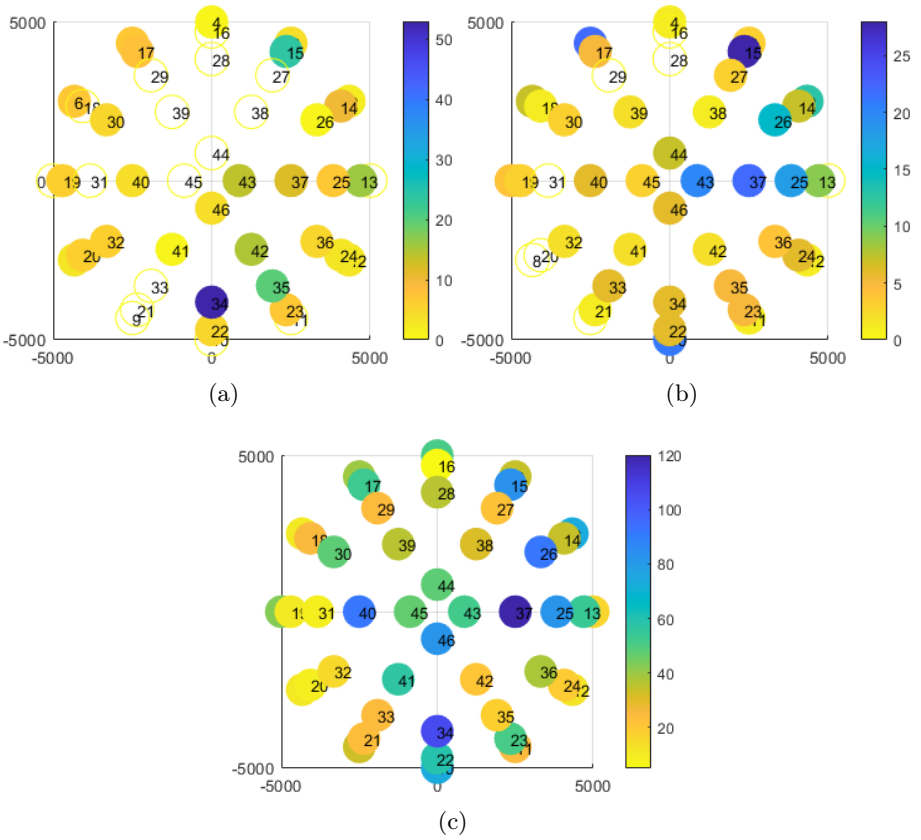
Rozkład błędów pomiędzy wszystkimi trzema akcjami niebezpiecznymi był różny dla obu dyscyplin. Zawodnicy Karate byli w równym stopniu myleni ze wszystkimi pozostałymi akcjami niebezpiecznymi (rys. 135). Pomyłki z akcją uderzenie dotyczyły 24 zawodników, a każdy z nich mylony był średnio z 11 różnych perspektyw. Dotyczyły przede wszystkim kopnięć wykonanych w tarczę. Do błędów nieco częściej dochodziło, gdy kamera znajdowała się na wprost zawodnika. Błędna klasyfikacja, jako kopnięcie niskie dotyczyła każdego zawodnika, w co najmniej jednym a średnio 9 punktach. Punkty, w których częściej dochodziło do pomyłek znajdowały się z lewej strony zawodnika. Ponownie większość nagrań błędnie klasyfikowanych zawierała kopnięcia w tarczę. Pomyłki z akcją kopnięcie wysokie dotyczyły praktycznie wszystkich zawodników. Każdy z nich mylony był średnio z 9 różnych perspektyw, z czego największa liczba w punktach znajdujących się z lewej strony z tyłu zawodnika. W przypadku pomyłek z tą akcją znacznie częściej dochodziło do niej, gdy zawodnik kopał w powietrze.

Zawodnicy Taekwondo z kolei byli myleni głównie z akcją kopnięcie wysokie boczne, chodź do pomyłek z pozostałymi akcjami niebezpiecznymi również dochodziło (rys. 136). Do częstszego mylenia kopnięcia wysokiego prostego z uderzeniem częściej dochodziło po lewej stronie zawodnika. Pomyłki te dotyczyły kopnięć, głównie w tarczę, wykonanych przez 20 różnych zawodników. Do pomyłek dochodziło średnio z 5 różnych perspektyw. Co ciekawe nagrania, któ-

re były mylone z jednej perspektywy z uderzeniem z innej (średnio 6 innych) mylone były z kopnięciem niskim. Co najmniej jedno kopnięcie wysokie proste każdego zawodnika Taekwondo, wykonane głównie w tarczę, mylone było z kopnięciem wysokim bocznym średnio dla 15 różnych perspektyw. Do błędów częściej dochodziło, gdy wirtualna kamera znajdowała się po lewej stronie zawodnika. Znalazł się również zawodnik, który niezależnie od położenia wirtualnej kamery zawsze był błędnie klasyfikowany. Sposób wykonania tego uderzenia, przez tego zawodnika odbiegał od pozostałych - uderzenie w tarczę wykonywane było znacznie wolniej i ostrożniej.



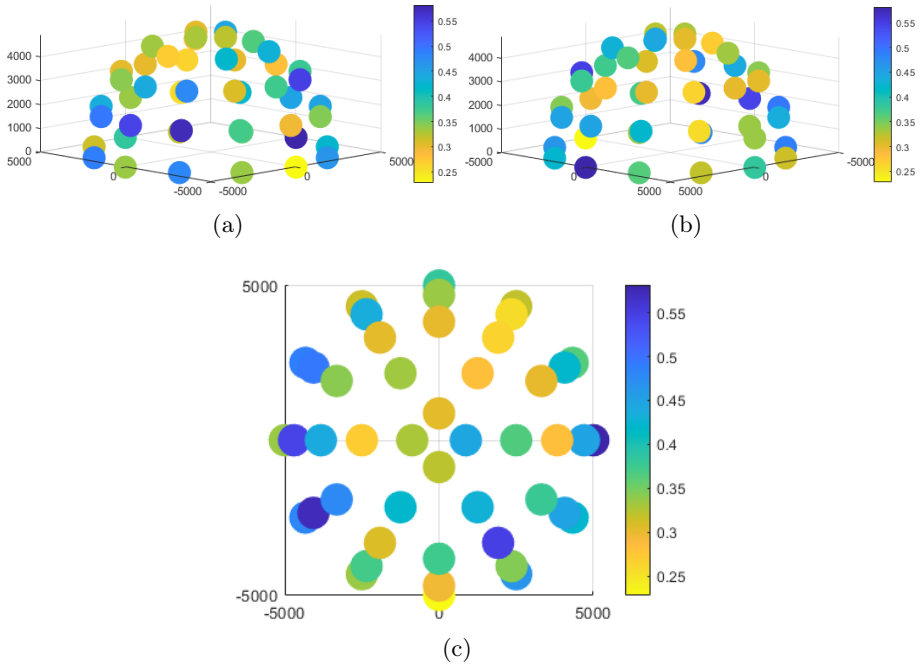
Rysunek 135: Kopuła liczby błędnych klasyfikacji wraz z ich numerami dla sieci CNN, w których akcja kopnięcie wysokie proste wykonana przez zawodników Karate mylona była z akcją: (a) uderzenie, (b) kopnięcie niskie, (c) kopnięcie wysokie boczne.



Rysunek 136: Kopuła liczby błędnych klasyfikacji wraz z ich numerami dla sieci CNN, w których akcja kopnięcie wysokie proste wykonana przez zawodników Taekwondo mylona była z akcją: (a) uderzenie, (b) kopnięcie niskie, (c) kopnięcie wysokie boczne.

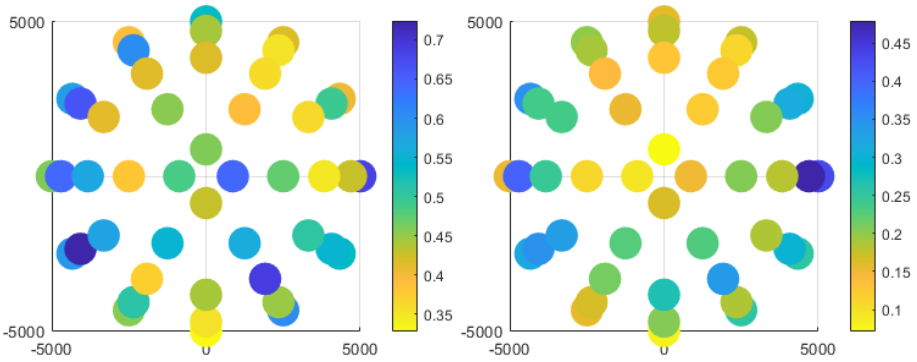
6.5.10 Akcja kopnięcie wysokie boczne

Czułość klasyfikacji ostatniej z rozpatrywanych akcji - kopnięcia wysokiego bocznego, podobnie jak w przypadku sieci LSTM w bardzo dużym, stopniu zależała od położenia wirtualnej kamery. Gdy kamera znajdowała się bliżej podłoża do pomyłek dochodziło znacznie częściej a czułość klasyfikacji spadała do 42%. Lepsze wyniki uzyskiwano, gdy kamera znajdowała za zawodnikiem z jego lewej strony i wynosiła do 77% (rys. 137).



Rysunek 137: Kopuła wartości PBA sieci CNN, dla akcji kopnięcie wysokie proste, widziana z trzech różnych perspektyw (a) z boku, (b) z boku, (c) od góry.

Tak samo jak w przypadku poprzednich eksperymentów, znacznie częściej do błędnej klasyfikacji dochodziło w przypadku zawodników Karate - czułość klasyfikacji wynosiła od 27% do 67%. Najniższe wyniki uzyskano, gdy wirtualna kamera była po prawej stronie zawodnika blisko podłoża (rys. 138 a). Wśród zawodników Taekwondo do większej liczby pomyłek dochodziło, gdy wirtualna kamera znajdowała się po lewej stronie zawodnika również blisko podłoża (rys. 138 b). Czułość klasyfikacji w przypadku tej grupy wahała się od 52% do 93%. Zestawienie poszczególnych rodzajów pomyłek znajduje się w tabelach 33 i 34.



(a) (b)

Rysunek 138: Kopuła wartości PBA sieci CNN, dla akcji kopnięcie wysokie boczne, widziana z góry dla zawodników (a) Karate, (b) Taekwondo.

Tabela 33: Lista punktów, dla których dochodziło do pomyłek akcji kopnięcie wysokie proste (na 29256 klasyfikacji), wśród zawodników Karate

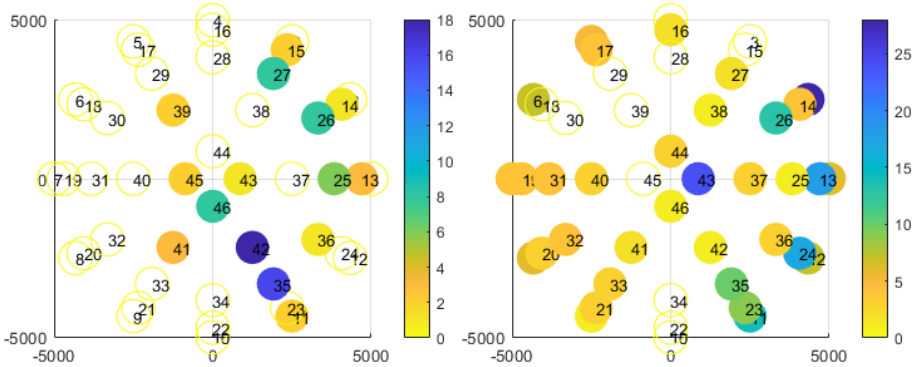
Akcja	Punkty	Pomyłki
Chód osób zdrowych	45	1
Chód osób chorych	11 13 14 15 25 26 27 35 36 39 41 42 43 45 46	81
Stanie	1 2 14 15 24 41 42 45 46	15
Obroty	1 2 3 4 12 13 14 15 19 27 28 29 31 32 34 37 38 39 40 41 42 43 44 46	133
Schylanie się	1 4 5 12 13 14 17 19 20 22 23 24 25 27 31 32 33 35 40 41 42	64
Uderzenie	Wszystkie	906
Kopnięcie niskie	Wszystkie	1407
Kopnięcie w. proste	Wszystkie	11439

Tabela 34: Lista punktów, dla których dochodziło do pomyłek akcji kopnięcie wysokie proste (na 32798 klasyfikacji), wśród zawodników Taekwondo

Akcja	Punkty	Pomyłki
Chód osób zdrowych	3 10 16 28 29 33 34 39 43	20
Chód osób chorych	Wszystkie oprócz 3 4 10 15 18 22 28 29 30... 34 39 45	220
Stanie	10 12 13 23 24 35 36 42 43 45 46	29
Obroty	Wszystkie oprócz 14 38 44	465
Schylanie się	18 34 41 45	8
Uderzenie	25 34 35 36 39 42	11
Kopnięcie niskie	2 6 10 11 14 15 32	17
Kopnięcie w. proste	Wszystkie	6375

Do błędnej klasyfikacji jako chód osób zdrowych dochodziło bardzo rzadko. W przypadku zawodników Karate do takiego błędu doszło tylko raz, a Taekwondo 20 razy dla 3 różnych zawodników. Nieco częściej, gdy po wykonaniu kopnięcia zawodnik dodatkowo się przemieszczał, jego ruch klasyfikowany był jako chód osób chorych. W przypadku zawodników Karate punkty w których do błędów dochodziło częściej znajdowały się po lewej stronie zawodnika. Podobnie w przypadku zawodników Taekwondo, chodź wśród nich zdarzały się pomyłki również dla innych położeń wirtualnej kamery (rys. 139).

Błędna klasyfikacja z akcją stanie zdarzała się bardzo rzadko, w przypadku kilku uderzeń w tarczę. Podobnie w przypadku pomyłek z akcją schylanie, przy czym dla zawodników Karate do tego rodzaju błędu dochodziło nieco częściej. Średnio do błędów dochodziło dla 2-3 różnych położeń wirtualnej kamery, zarówno wśród zawodników Karate jak i Taekwondo. Przy czym liczba zawodników Karate, których kopnięcia zostały pomylone ze staniem była większa.

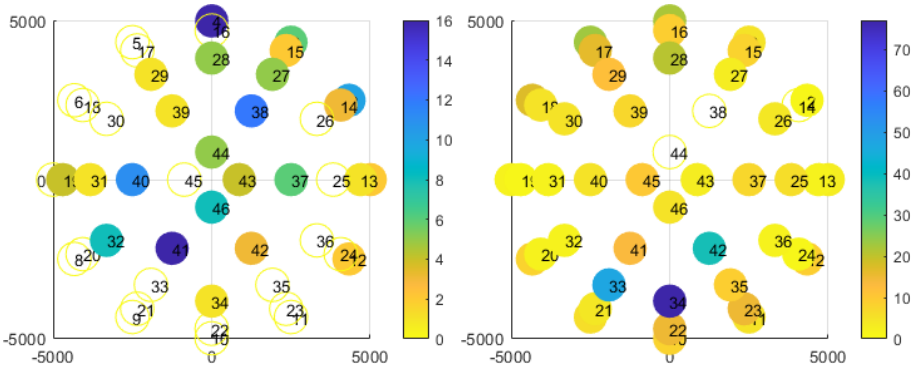


(a) (b)

Rysunek 139: Kopuła liczby błędnych klasyfikacji wraz z ich numerami dla sieci CNN, w których akcja kopnięcie wysokie boczne mylona była z akcją chód osób chorych dla zawodników: (a) Karate, (b) Taekwondo.

Znacznie częściej do błędów dochodziło z akcją obroty. Kopnięcia 19 zawodni-

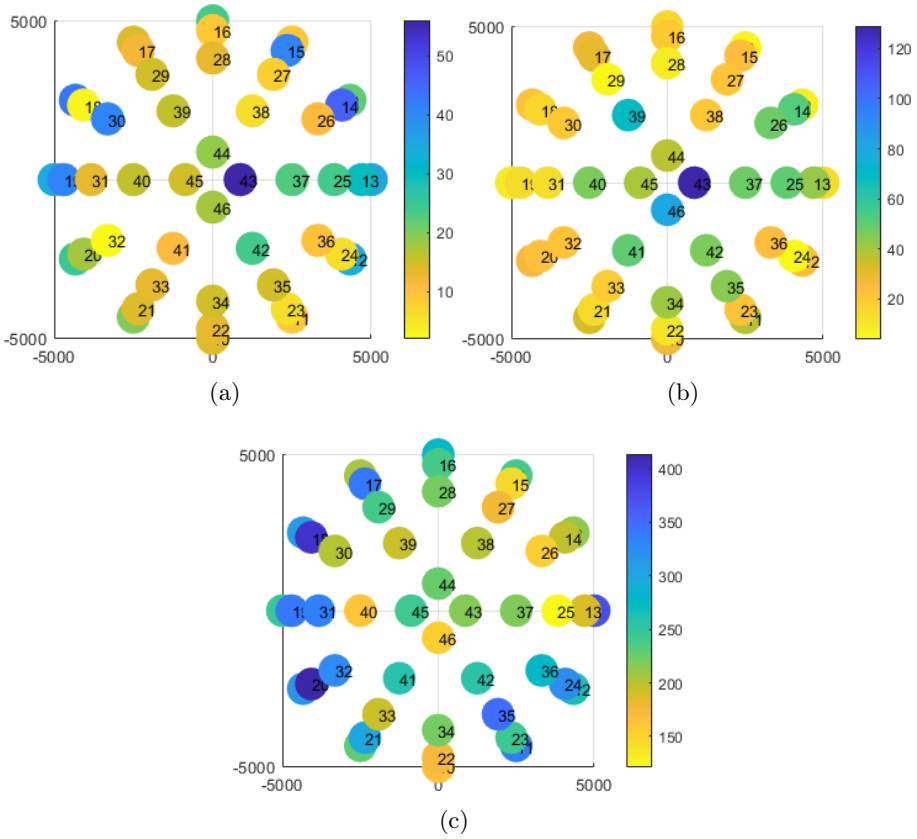
ków Karate widziane średnio z 3 różnych perspektyw były błędnie klasyfikowane jako obrót. Większość punktów w których dochodziło do pomyłek znajdowała się za zawodnikiem (rys. 140 a). W przypadku zawodników Taekwondo błędy dotyczyły aż 22 z nich w średnio 8 różnych punktach. Największą liczbę błędów uzyskano gdy wirtualna kamera znajdowała się tuż przed zawodnikiem (rys. 140 b).



(a) (b)

Rysunek 140: Kopuła liczby błędnych klasyfikacji wraz z ich numerami dla sieci CNN, w których akcja kopnięcie wysokie boczne mylona była z akcją obroty dla zawodników: (a) Karate, (b) Taekwondo.

Podobnie jak w przypadku kopnięcia wysokiego prostego, rozkład błędnych klasyfikacji z pozostałymi akcjami niebezpiecznymi jest skrajnie różny dla zawodników obu drużyn. Kopnięcia 22 różnych zawodników Karate były błędnie klasyfikowane jako uderzenie średnio dla 14 różnych położeń wirtualnej kamery. Do pomyłek z tą akcją częściej dochodziło gdy wirtualna kamera znajdowała się po bokach zawodnika (rys. 141 a). Błędy dotyczyły nagrań w których zawodnik kopał zarówno w tarczę jak i w powietrze. Pomyłki z kopnięciem niskim dotyczyły każdego zawodnika widzianego z co najmniej jednej perspektywy (średnio 11). W punktach znajdujących się bliżej szczytu kopuły znacznie częściej dochodziło do błędów (rys. 141 b). Zdecydowana większość nagrań, które zostały nieprawidłowo sklasyfikowane zawierała kopnięcia w tarczę. Nagrania z kopnięciem w powietrze częściej natomiast były klasyfikowane, jako kopnięcie wysokie proste. Błędy dotyczyły wszystkich zawodników, średnio w aż 35 różnych położeniach wirtualnej kamery. Znacznie częściej do błędnych klasyfikacji dochodziło, gdy kamera znajdowała się po prawej stronie zawodnika (rys. 141 c).

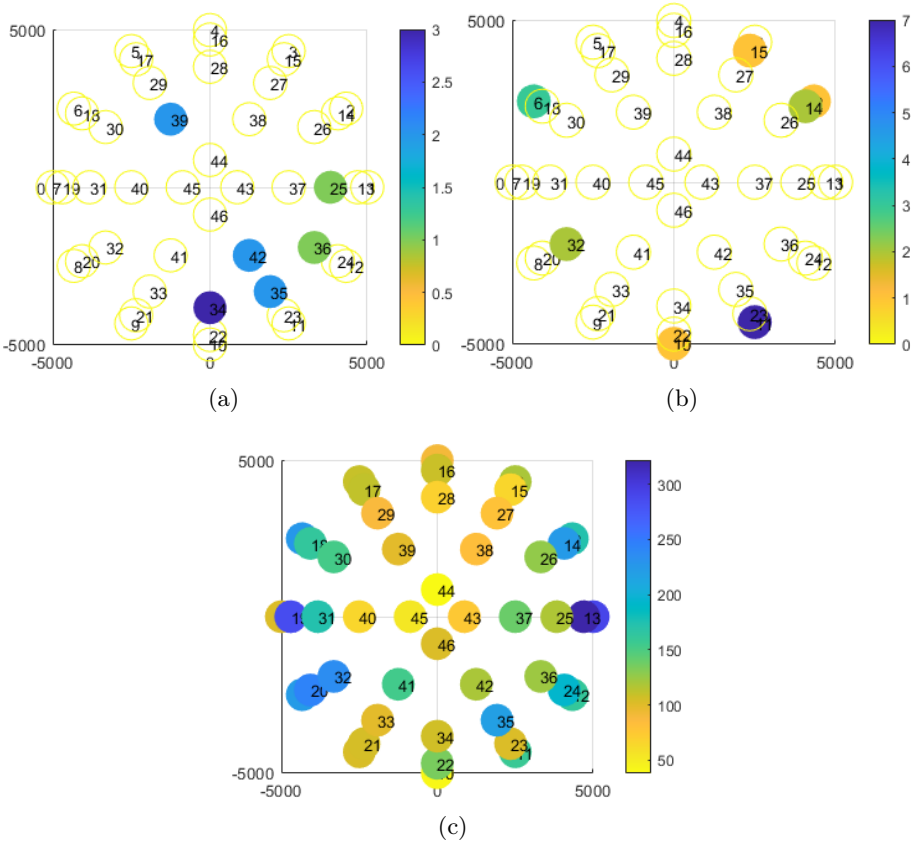


(a) (b)

(c)

Rysunek 141: Kopuła liczby błędnych klasyfikacji wraz z ich numerami dla sieci CNN, w których akcja kopnięcie wysokie boczne wykonana przez zawodników Karate mylona była z akcją: (a) uderzenie, (b) kopnięcie niskie, (c) kopnięcie wysokie boczne.

Wśród zawodników Taekwondo pomyłki z akcjami uderzenie i kopnięcie niskie, zdarzały się bardzo sporadycznie. Jeśli już do błędu doszło, dotyczył on najczęściej 1/2 kopnięć w tarczę danego zawodnika widzianego z jednej perspektywy. Podobnie jak w przypadku zawodników Karate, do pomyłek z akcją kopnięcie wysokie proste dochodziło dla każdego zawodnika Taekwondo widzianego średno z 26 różnych perspektyw. Do błędów znacznie częściej dochodziło, gdy wirtualna kamera znajdowała się po bokach zawodnika (rys. 142). Ponownie jak w przypadku poprzednich eksperymentów, błędy dotyczyły głównie nagrań, w których zawodnik uderzał w tarczę.



(a) (b)

(c)

Rysunek 142: Kopuła liczby błędnych klasyfikacji wraz z ich numerami dla sieci CNN, w których akcja kopnięcie wysokie boczne wykonana przez zawodników Taekwondo mylona była z akcją: (a) uderzenie, (b) kopnięcie niskie, (c) kopnięcie wysokie boczne.

6.5.11 Wnioski

Podobnie jak w przypadku sieci LSTM, w przypadku pewnych akcji widać ogromny wpływ rzutowania perspektywicznego, na czułość ich klasyfikacji. W zależności od ustawienia wirtualnej kamery to samo nagranie klasyfikowane było poprawnie, lub mylone, z co najmniej jedną inną akcją. Jest to szczególnie widoczne dla wszystkich trzech rodzajów kopnięć.

Sieć CNN ponownie poradziła sobie najgorzej z poprawną klasyfikacją osób zdrowych. Odbyło się to kosztem stosunkowo wysokiej, czułości klasyfikacji osób chorych. Wśród osób chorych można zauważyć wyraźny wpływ danego schorzenia, na czułość klasyfikacji oraz na rodzaj innych akcji, z którymi dochodziło do błędów. Przykładowo chód pacjentów ze zwyrodnieniem kręgosłupa mylony był tylko z chodem osób zdrowych, podczas gdy osoby z innymi schorzeniami myleni byli z pozostałymi akcjami.

Akcje statyczne, za wyjątkiem schylania się, cechowały się bardzo wysoką rozpoznawalnością. W ich przypadku również widać wyraźny wpływ położenia wirtualnej kamery zarówno, na czułość klasyfikacji jak i rodzaje pomyłek.

Na czułość klasyfikacji akcji potencjalnie niebezpiecznych wpływ miały trzy rzeczy: położenie wirtualnej kamery, rodzaj dyscypliny sportowej jaką uprawiał dany zawodnik, oraz to czy uderzenie/kopnięcie wykonywane było w tarczę czy w powietrze. Rodzaj akcji, z którą dochodziło do pomyłek w dużej mierze zależał właśnie od sposobu wykonywania danej techniki.

6.6 Porównanie wyników dla obu rodzajów sieci

Biorąc pod uwagę tylko dokładność klasyfikacji wśród omawianych sieci, sieci LSTM uzyskały znacznie lepsze rezultaty niż sieci CNN. Jednakże podobnie jak w przypadku danych trójwymiarowych, na dokładność klasyfikacji dominujący wpływ miała jedna z wybranych akcji - chód osób zdrowych. Pomijając tą akcję, dokładność klasyfikacji sieci CNN jest znacznie wyższa niż LSTM.

W przypadku obu rodzajów sieci można zauważyć wpływ rzutowania perspektywicznego zarówno, na czułość klasyfikacji poszczególnych akcji, jak i na typ akcji, z którą dochodzi do pomyłek. Dodatkowo niezależnie od omawianej akcji, błędy najczęściej dotyczą tych samych osób, jednakże w różnym stopniu. Przykładowo sieć CNN, pomyli 3 z 5 uderzeń danej osoby widziane z danej perspektywy, a sieć LSTM 4 z 5. Przy czym pomyłki te mogą być podobne (klasyfikacja uderzenia, jako stanie przez obie sieci) lub różne (jedna sieć dane uderzenie sklasyfikuje, jako stanie a druga, jako obrót). Bardziej szczegółowe porównanie uzyskanych wyników przeprowadzono z podziałem na typy akcji: chody, akcje statyczne, oraz akcje potencjalnie niebezpieczne.

6.6.1 Akcje - chody

Czułość klasyfikacji akcji chód osób zdrowych była dużo wyższa dla sieci LSTM. Dodatkowo w przypadku tych sieci można zaobserwować bardzo duży wpływ rzutowania perspektywicznego na poprawność klasyfikacji. Lepsze rezultaty uzyskiwano, gdy wirtualna kamera znajdowała się przed nagrywaną osobą, lub z jej lewej strony. Sieć CNN dla większości położeń wirtualnej kamery uzyskiwała podobne, niskie, rezultaty. W momencie, gdy wirtualna kamera znajdowała się po lewej stronie osoby, czułość klasyfikacji była najniższa.

Akcja ta była błędnie klasyfikowana głównie, jako akcja chód osób chorych. Pomyłki z innymi akcjami również się zdarzały, jednakże były to pojedyncze przypadki. W przypadku obu sieci, znacznie częściej dochodziło do błędów, gdy dana osoba nagrywana była za pomocą oprogramowania Vicon Nexus.

Drugi rodzaj chodu - chód osób chorych, osiągał znacznie lepsze rezultaty przy wykorzystaniu sieci CNN. Sieci te, były bardziej wrażliwe na wszelkie nieprawidłowości w chodzie, dzięki czemu lepiej rozpoznawały tę akcję. Odbyło się to kosztem znacznie gorszej, czułości klasyfikacji dla akcji chód osób zdrowych. W przypadku obu rodzajów sieci, można zauważyć spadek, czułości klasyfikacji gdy kamera znajduje się stosunkowo blisko podłoża. Dodatkowo, ponownie widać bardzo znaczący wpływ danego schorzenia nie tylko na czułość klasyfikacji, ale też częstotliwość pomyłek z daną akcją.

Pacjenci, którzy byli myleni z przez sieć CNN byli również myleni przez sieć LSTM. Punkty, w których dochodziło do danego błędu czasem się pokrywały. Zdarzały się też przypadki, gdy jedna sieć błędnie zaklasyfikowała dane przejście widziane z danej perspektywy, jako stanie a druga, jako obrót. Dodatkowo w przypadku sieci CNN sporadyczne pomyłki z innymi akcjami nie tylko zdarzały się rzadziej, ale i w przypadku pewnych schorzeń w ogóle do nich nie dochodziło.

W przypadku obu rodzajów sieci, pacjenci ze zwyrodnieniem kręgosłupa byli grupą nie tylko z najwyższą, czułością klasyfikacji, ale też poza dwoma przypadkami dla sieci LSTM myleni byli tylko z chodem osób chorych. Może to wynikać z faktu, iż bóle kręgosłupa w odcinku lędźwiowym wpływają na usztywnienie pacjenta, ale nie powodują żadnych dodatkowych ruchów dłoni, czy też nie wpływają zbytnio na przebieg akcji. W przypadku pozostałych pacjentów, w zależności od schorzenia dochodzi do szeregu dodatkowych nieprawidłowości. U pacjentów z chorobą Parkinsona poza drżeniem dłoni dochodzi do tzw. zamrożeń chodu. Podobnie osoby po udarze czy po endoprotezoplastyce mają bardziej ograniczoną ruchomość kończyn dolnych, co w pewnych przypadkach może powodować błędne klasyfikacje, jako inne akcje.

Pomyłki z chodem osób zdrowych, ponownie dotoczył pacjentów w znacznie lepszej kondycji. Osoby po długiej rehabilitacji, oraz tacy, u których objawy choroby były znikome, były częściej klasyfikowane, jako chód osób zdrowych. Co warte zaznaczenia, wpływ rzutowania perspektywicznego był różny dla danego schorzenia. Pacjenci po endoprotezoplastyce stawu biodrowego byli nieprawidłowo klasyfikowani, jako osoby zdrowe przez sieć CNN praktycznie z każdej perspektywy poza punktem znajdującym się na wprost pacjenta. W przypadku sieci LSTM można zauważyć wyraźny podział kopuły na pół - punkty po prawej stronie pacjenta uzyskiwały znacznie gorsze rezultaty niż te po jego lewej. Podobnie jak dla pacjentów po udarze czy z chorobą Parkinsona - sieć LSTM znacznie lepiej klasyfikowała tę akcje, gdy kamera znajdowała się po lewej stronie pacjenta. W przypadku sieci CNN natomiast lepsze rezultaty osiągały punkty znajdujące się przed lub za pacjentem. U osób ze zwyrodnieniem kręgosłupa nie widać aż tak dużego wpływu rzutowania jak w przypadku pozostałych schorzeń, niezależnie od rodzaju sieci.

6.6.2 Akcje statyczne

Czułość klasyfikacji akcji stanie w przypadku obu rodzajów sieci była bardzo wysoka, z nieznaczną przewagą sieci CNN. W przypadku obu sieci akcja ta jest nieco gorzej klasyfikowana, gdy wirtualna kamera znajduje się tuż nad podło-

żem. Przy czym dla sieci LSTM wpływ położenia wirtualnej kamery, na czułość klasyfikacji jest znacznie większy. W przypadku obu sieci do błędnej klasyfikacji dochodziło głównie z akcjami obroty, uderzenie lub schylanie się. Przy czym do pomyłek z akcjami obroty i uderzenie dochodziło dla każdego położenia wirtualnej kamery. Pomyłki z pozostałymi akcjami były sporadyczne. Najczęściej dotyczyły tych samych osób i ich konkretnych nagrań.

Wpływ rzutowania perspektywicznego na akcję obroty ponownie był podobny w przypadku obu rodzajów sieci - punkty znajdujące się po prawej stronie osoby uzyskiwały znacznie lepsze rezultaty. W przypadku sieci CNN, czułość klasyfikacji była nieznacznie wyższa niż dla sieci LSTM. Do pomyłek najczęściej dochodziło z akcją stanie. Sieci LSTM błędnie rozpoznawały tę akcje również, jako uderzenie lub kopnięcie wysokie proste a CNN tylko kopnięcie wysokie proste. Pomyłki z pozostałymi akcjami były bardziej sporadyczne, aczkolwiek się zdarzały. Podobnie błędy dotyczyły tych samych osób w różnym stopniu.

Akcja schylanie się w odróżnieniu od dwóch wcześniejszych, była lepiej klasyfikowana przez sieć LSTM. Dodatkowo zależność, pomiędzy czułością klasyfikacji a położeniem wirtualnej kamery jest różna dla obu sieci. Sieć LSTM uzyskiwała lepsze rezultaty dla punktów znajdujących się bliżej podłoża, a sieć CNN odwrotnie - wraz ze zbliżaniem się wirtualnej kamery do szczytu kopuły, czułość klasyfikacji rosła. Pomyłki obejmowały wszystkie akcje, najczęściej były to akcje stanie, uderzenie czy kopnięcie wysokie proste oraz obroty. Przy czym, w przypadku obu sieci dochodziło też do znaczącej liczby pomyłek z akcjami chód osób zdrowych lub chorych. Zdecydowana większość pomyłek dotyczyła nagrań, w których skłony wykonywane były dynamicznie (bardzo krótkie nagrania), lub odwrotnie dana osoba w skłonie przebywała dość długo i wykonywała go powolnie (klasyfikacja, jako stanie).

6.6.3 Akcje potencjalnie niebezpieczne

W przypadku wszystkich czterech potencjalnie niebezpiecznych akcji można zauważyć mniejszy lub większy wpływ rzutowania perspektywicznego. Dodatkowo wpływ, na czułość klasyfikacji miała dyscyplina sportowa uprawiana przez daną osobą. Z kolei na to, z jaką akcją dochodziło do pomyłek największy wpływ miał rodzaj uderzenia - w powietrze albo w tarczę. W zależności od akcji, ustawienie celu albo jego brak powodowały różnego rodzaju odchylenia od wzorca.

W przypadku akcji uderzenie, ponownie lepsze rezultaty uzyskały sieci CNN. W przypadku tego rodzaju sieci wpływ rzutowania perspektywicznego, na czułość klasyfikacji jest niewielki. Nieznacznie wyższe rezultaty uzyskały punkty znajdujące się po lewej stronie zawodnika. Sieci LSTM natomiast, ponownie uzyskiwały gorsze rezultaty, gdy wirtualna kamera znajdowała się z lewej strony zawodnika. Co warte zaznaczenia w przypadku obu sieci, wpływ rzutowania perspektywicznego jest inny dla zawodników danej dyscypliny sportowej. Zawodnicy Karate byli lepiej klasyfikowani przez sieć LSTM, gdy wirtualna kamera znajdowała się tuż przed lub za nimi. Zawodnicy Taekwondo natomiast byli dość dobrze klasyfikowani z większości perspektyw, z lekką przewagą punktów po prawej stronie zawodnika. Wyjątek stanowiło kilka punktów przed zawodnikiem. W przypadku sieci CNN wpływ rzutowania perspektywicznego, na czułość klasyfikacji widać tylko dla zawodników Taekwondo - punkty za zawodnikiem z jego lewej strony, oraz przed po prawej osiągają nieco gorsze rezultaty.

Wśród zawodników Karate dla obu rodzajów sieci najczęściej dochodziło do

błędnej klasyfikacji uderzeń, jako kopnięcia wysokiego prostego, bocznego lub uderzenia, nieco rzadziej, jako obrotu lub stania (częściej w przypadku sieci CNN). Pomyłki wśród zawodników Taekwondo zdarzały się znacznie rzadziej i dotyczyły głównie akcji kopnięcie wysokie proste, stanie, obroty i w przypadku sieci LSTM schylanie się. Co warte zaznaczenia w przypadku obu rodzajów sieci do pomyłek z akcją stanie znacznie częściej dochodziło, gdy wirtualna kamera znajdowała się z prawej strony zawodnika. Dotyczyło to przede wszystkim nagrań zawierających uderzenie w tarczę.

Kopnięcie niskie spośród wszystkich akcji niebezpiecznych charakteryzowało się najgorszą, czułością klasyfikacji, ponownie nieco lepszą dla sieci CNN. Wpływ rzutowania perspektywicznego jest bardzo widoczny. W przypadku obu sieci punkty znajdujące się bliżej podłoża osiągają znacznie lepsze, przekraczające 70% rezultaty, podczas gdy punkty znajdujące się bliżej szczytu kopuł uzyskiwały znacznie gorsze wyniki sięgające tylko 20% na jej szczycie. Dodatkowo sieci LSTM, ponownie osiągały gorsze rezultaty, gdy wirtualna kamera znajdowała się po lewej stronie zawodnika. Zdecydowana większość pomyłek dotyczyła wszystkich trzech pozostałych akcji niebezpiecznych ze zdecydowaną przewagą kopnięcia wysokiego prostego. W przypadku obu rodzajów sieci do pomyłek z akcją uderzenie rzadziej dochodziło, gdy wirtualna kamera znajdowała się za zawodnikiem. Natomiast do błędnej klasyfikacji z kopnięciem wysokim bocznym dochodziło, ponownie w przypadku obu rodzajów sieci, najczęściej, gdy wirtualna kamera znajdowała się na szczycie kopuły.

Czułość klasyfikacji akcji kopnięcie wysokie proste również była wyższa dla sieci CNN. W przypadku obu rodzajów sieci, wyższe wyniki uzyskiwano, gdy wirtualna kamera znajdowała się za zawodnikiem lub z jego prawej strony. Czułość klasyfikacji dla zawodników poszczególnych grup zależała od rodzaju sieci - dla sieci LSTM była ona wyższa dla zawodników Taekwondo, a dla sieci CNN odwrotnie lepszą czułość klasyfikacji uzyskiwali zawodnicy Karate. Wpływ rzutowania perspektywicznego, na czułość klasyfikacji dla zawodników Taekwondo jest podobny dla obu rodzajów sieci - lepsze rezultaty osiągają punkty znajdujące się po prawej stronie zawodnika, stosunkowo blisko podłoża. W przypadku zawodników Karate zależność ta jest znacznie mniej wyraźna, przy czym obie sieci uzyskują gorsze rezultaty, gdy wirtualna kamera jest bliżej szczytu kopuły. Kopnięcia wysokie proste zawodników Karate są częściej mylone z kopnięciem niskim, przez oba rodzaje sieci. Kopnięcia zawodników Taekwondo natomiast są częściej mylone z kopnięciem wysokim bocznym. W przypadku obu sieci, niezależnie od dyscypliny sportowej uprawianej przez zawodnika dochodzi też do pomyłek z pozostałymi akcjami z przewagą uderzenia i obrotu.

Kopnięcie wysokie boczne, w odróżnieniu do danych trójwymiarowych, cechowało się wysoką częstotliwością błędów, nieco wyższą w przypadku sieci LSTM. Wpływ rzutowania perspektywicznego, na czułość klasyfikacji jest odwrotny dla obu sieci - sieci LSTM uzyskiwały lepsze rezultaty dla punktów położonych bliżej podłoża, a sieci CNN dla punktów bliżej szczytu kopuły. Podobieństwa we wpływie rzutowania perspektywicznego są natomiast widoczne dla zawodników Taekwondo - w przypadku obu rodzajów sieci punkty bliżej szczytu kopuły uzyskiwały lepsze rezultaty. Ten rodzaj kopnięcia był błędnie klasyfikowany najczęściej, jako jedna z pozostałych akcji niebezpiecznych. Pomyłki wśród zawodników Karate najczęściej zdarzały się z akcjami kopniecie wysokie proste, kopnięcie niskie bądź uderzenie. W przypadku zawodników Taekwondo natomiast pomyłki dotyczyły głównie kopnięcie wysokie proste, obroty

i chód osób chorych. Bardzo częste pomyłki zawodników Karate z kopnięciem niskim, oraz fakt, że do tego rodzaju pomyłek rzadko dochodziło wśród zawodników Taekwondo mogą wynikać z faktu, iż kopnięcie niskie wykonywali tylko zawodnicy Karate.

7 Podsumowanie

Przygotowana na potrzeby zaplanowanych eksperymentów baza danych okazała się wystarczająco duża by nie doszło do przeuczenia głębokich sieci neuronowych. Pozwoliła ona na wykazanie wpływu rzutowania perspektywicznego, na jakość klasyfikacji przy użyciu dwóch różnych rodzajów głębokich sieci neuronowych. Wpływ ten jest różny dla każdego typu akcji. Dodatkowo, można zauważyć korelację pomiędzy ustawieniem wirtualnej kamery, a rodzajem akcji, z którą dochodzi do pomyłek.

Dodatkowo, co zaskakujące wpływ rzutowania perspektywicznego nie był symetryczny. Punkty leżące po przeciwnych stronach kopuły potrafiły uzyskać skrajnie różne rezultaty. W przypadku sieci LSTM zdecydowanie gorsze wyniki osiągano, gdy wirtualna kamera znajdowała się po lewej stronie osoby (po prawej stronie kopuły widzianej z góry).

Szczegółowa analiza wyników wykazała, iż, akcją, która uzyskała najgorsze rezultaty był chód osób zdrowych. Można zaobserwować tendencję (większą w przypadku sieci CNN) do bardzo krytycznej oceny ruchu, przez co wszelkie nieprawidłowości i odstępstwa od wzorca powodują klasyfikacje chodu osoby zdrowej, jako chorej. Dodatkowo w zależności od schorzenia danego pacjenta i perspektywy, z jakiej jest on widoczny do pomyłek z akcją chód osób zdrowych dochodzi w różnym stopniu. Wszystkie omawiane schorzenia w różnym stopniu wpływają na sposób przemieszczania się. Dlatego też, na potrzeby przyszłych eksperymentów akcje te zostaną zgrupowane w jedną akcję chód.

Jednakże, wykrywanie stanu zdrowia danej osoby może być użyteczne z punktu widzenia bezpieczeństwa. Dlatego też, w przyszłości rozważane jest przeprowadzenie dodatkowych eksperymentów skupiających się tylko na chodzie, w ramach, których zostaną utworzone klasyfikatory mające na celu rozpoznanie danego schorzenia na podstawie samego chodu.

Poza rzutowaniem perspektywicznym, wpływ, na jakość klasyfikacji miały również inne czynniki. W przypadku akcji niebezpiecznych był to między innymi sposób wykonania danego uderzenia/kopnięcia. Można zaobserwować wyraźną różnicę, w jakości klasyfikacji przy uderzeniach w tarczę i w powietrze. Dlatego też, należy rozważyć czy w przyszłości nie podzielić tych nagrań na osobne kategorie, w których uderzenia/kopnięcia w powietrze traktowane będą, jako wymachy, a w tarczę, jako faktycznie akcje niebezpieczne.

Dodatkowo w przypadku niektórych akcji (np. stanie), zdecydowana większość wykonywała ją w konkretny sposób. Gdy znalazła się osoba, która wychodziła poza ten schematy (stała i machała rękami), jej nagrania był błędnie klasyfikowane przez obie sieci. Dlatego też, niezbędne jest dalsze rozszerzanie omawianej bazy danych. Można to zrobić na dwa sposoby. Pierwszy, bardziej czasochłonny, zakłada wykonanie większej liczby różnorodnych nagrań. Drugi zakłada zmodyfikowanie już istniejących. Przy czym zmiany te nie mogą zachodzić w przypadku pojedynczych trajektorii, ale całych segmentów ciała, np. dodanie wymachów rąk podczas chodu, czy dodanie przemieszczania się podczas uderzenia.

Ponieważ celem niniejszej pracy jest ogólna klasyfikacja zachowań postaci ludzkiej, oraz spraw dzenie jak w zależności od ustawień kamery zmienia się rozpoznawalność po szczególnych akcji, do rozpoznawania wybrano tylko podstawowe wersje sieci- jednowymiarowy CNN, oraz wspomniane dwa warianty

sieci LSTM. Wprowadzenie bardziej złożonych modyfikacji pozostaje tematem do rozważenia w przyszłości.

Algorytm podziału na podzbiory

Przygotowanie bazy 2D

Metoda wizualizacji

wprowadzenie miar

wstawić: skopiowane z tekstu wyjaśnienia dlaczego ograniczono się do LSTM i CNN, wyjaśnienia o augmentacji i wpływie zakłóceń jako badania na przyszłość, dopisać: badanie wpływu parametrów wewnętrznych głównie ogniskowej kamery wirtualnej, wpływu rozmieszczenia punktów/markerów na aktorze na jakość klasyfikacji

Dalsze badania: Zagadnienie to wymaga większych przemyśleń i może być podstawą do dalszych prac i badań. Dlatego też zrezygnowano z ich wprowadzania do danych.

Warto natomiast rozważyć analizę wpływu poszczególnych zakłóceń, na jakość rozpoznawania poszczególnych akcji, jednakże nie jest to celem niniejszej pracy.

Literatura

[1] Jake K Aggarwal and Michael S Ryoo. Human activity analysis: A review. *Acm Computing Surveys (Csur)*, 43(3):1–43, 2011.

[2] Maryam Ziaeefard and Robert Bergevin. Semantic human activity recognition: A literature review. *Pattern Recognition*, 48(8):2329–2345, 2015.

[3] Sarvesh Vishwakarma and Anupam Agrawal. A survey on activity recognition and behavior understanding in video surveillance. *The Visual Computer*, 29:983–1009, 2013.

[4] Wei Niu, Jiao Long, Dan Han, and Yuan-Fang Wang. Human activity detection and recognition for video surveillance. In *2004 IEEE international conference on multimedia and expo (ICME)(IEEE Cat. No. 04TH8763)*, volume 1, pages 719–722. IEEE, 2004.

[5] Yan Wang, Shuang Cang, and Hongnian Yu. A survey on wearable sensor modality centred human activity recognition in health care. *Expert Systems with Applications*, 137:167–190, 2019.

[6] Xiaokang Zhou, Wei Liang, I Kevin, Kai Wang, Hao Wang, Laurence T Yang, and Qun Jin. Deep-learning-enhanced human activity recognition for internet of healthcare things. *IEEE Internet of Things Journal*, 7(7):6429–6438, 2020.

[7] Han Sun and Yu Chen. Real-time elderly monitoring for senior safety by lightweight human action recognition. In *2022 IEEE 16th International Symposium on Medical Information and Communication Technology (ISMICT)*, pages 1–6. IEEE, 2022.

[8] Alina Roitberg, Alexander Perzylo, Nikhil Somani, Manuel Giuliani, Markus Rickert, and Alois Knoll. Human activity recognition in the context of

industrial human-robot interaction. In *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific*, pages 1–10. IEEE, 2014.

[9] Isidoros Rodomagoulakis, Nikolaos Kardaris, Vassilis Pitsikalis, E Mavroudi, Athanasios Katsamanis, Antigoni Tsiami, and Petros Maragos. Multimodal human action recognition in assistive human-robot interaction. In *2016 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pages 2702–2706. IEEE, 2016.

[10] Athanasios Anagnostis, Lefteris Benos, Dimitrios Tsaopoulos, Aristotelis Tagarakis, Naoum Tsolakis, and Dionysis Bochtis. Human activity recognition through recurrent neural networks for human–robot interaction in agriculture. *Applied Sciences*, 11(5):2188, 2021.

[11] Emmanuel Ramasso, Costas Panagiotakis, Denis Pellerin, and Michèle Rombaut. Human action recognition in videos based on the transferable belief model: application to athletics jumps. *Pattern analysis and Applications*, 11:1–19, 2008.

[12] Magdalena Pawlyta, Marek Hermansa, Agnieszka Szczęsna, Mateusz Janiak, and Konrad Wojciechowski. Deep recurrent neural networks for human activity recognition during skiing. In *Man-Machine Interactions 6: 6th International Conference on Man-Machine Interactions, ICMMI 2019, Cracow, Poland, October 2-3, 2019*, pages 136–145. Springer, 2020.

[13] Kristina Host and Marina Ivašić-Kos. An overview of human action recognition in sports based on computer vision. *Heliyon*, 8(6), 2022.

[14] Anandarup Mukherjee, Sudip Misra, P Mangrulkar, Muttukrishnan Rajarajan, and Yogachandran Rahulamathavan. Smartarm: A smartphone-based group activity recognition and monitoring scheme for military applications. In *2017 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS)*, pages 1–6. IEEE, 2017.

[15] Dariu M Gavrila. The visual analysis of human movement: A survey. *Computer vision and image understanding*, 73(1):82–98, 1999.

[16] Daniel Weinland, Remi Ronfard, and Edmond Boyer. A survey of vision-based methods for action representation, segmentation and recognition. *Computer vision and image understanding*, 115(2):224–241, 2011.

[17] T Subetha and S Chitrakala. A survey on human activity recognition from videos. In *2016 international conference on information communication and embedded systems (ICICES)*, pages 1–7. IEEE, 2016.

[18] Farhood Negin and François Bremond. Human action recognition in videos: A survey. *INRIA Technical Report*, 2016.

[19] Leonardo Onofri, Paolo Soda, Mykola Pechenizkiy, and Giulio Iannello. A survey on using domain and contextual knowledge for human activity recognition in video streams. *Expert Systems with Applications*, 63:97–111, 2016.

[20] Samitha Herath, Mehrtash Harandi, and Fatih Porikli. Going deeper into action recognition: A survey. *Image and vision computing*, 60:4–21, 2017.

[21] Di Wu, Nabin Sharma, and Michael Blumenstein. Recent advances in video-based human action recognition using deep learning: A review. In *2017 International joint conference on neural networks (IJCNN)*, pages 2865–2872. IEEE, 2017.

[22] Hong-Bo Zhang, Yi-Xiang Zhang, Bineng Zhong, Qing Lei, Lijie Yang, Ji-Xiang Du, and Duan-Sheng Chen. A comprehensive survey of vision-based human action recognition methods. *Sensors*, 19(5):1005, 2019.

[23] Yu Kong and Yun Fu. Human action recognition and prediction: A survey. *International Journal of Computer Vision*, 130(5):1366–1401, 2022.

[24] Thomas B Moeslund, Adrian Hilton, and Volker Krüger. A survey of advances in vision-based human motion capture and analysis. *Computer vision and image understanding*, 104(2-3):90–126, 2006.

[25] Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1798–1828, 2013.

[26] Alexander Toshev and Christian Szegedy. Deeppose: Human pose estimation via deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1653–1660, 2014.

[27] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7291–7299, 2017.

[28] Hao-Shu Fang, Shuqin Xie, Yu-Wing Tai, and Cewu Lu. Rmpe: Regional multi-person pose estimation. In *Proceedings of the IEEE international conference on computer vision*, pages 2334–2343, 2017.

[29] Leonid Pishchulin, Eldar Insafutdinov, Siyu Tang, Bjoern Andres, Mykhaylo Andriluka, Peter V Gehler, and Bernt Schiele. Deepcut: Joint subset partition and labeling for multi person pose estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4929–4937, 2016.

[30] Ke Sun, Bin Xiao, Dong Liu, and Jingdong Wang. Deep high-resolution representation learning for human pose estimation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5693–5703, 2019.

[31] GJ Burghouts, K Schutte, R J-M ten Hove, SP van den Broek, J Baan, O Rajadell, JR van Huis, J van Rest, P Hanckmann, H Bouma, et al. Instantaneous threat detection based on a semantic representation of activities, zones and trajectories. *Signal, Image and Video Processing*, 8:191–200, 2014.

[32] Heng Wang and Cordelia Schmid. Action recognition with improved trajectories. In *Proceedings of the IEEE international conference on computer vision*, pages 3551–3558, 2013.

[33] Haiam A Abdul-Azim and Elsayed E Hemayed. Human action recognition using trajectory-based representation. *Egyptian Informatics Journal*, 16(2):187–198, 2015.

[34] Ross Messing, Chris Pal, and Henry Kautz. Activity recognition using the velocity histories of tracked keypoints. In *2009 IEEE 12th international conference on computer vision*, pages 104–111. IEEE, 2009.

[35] Pyry Matikainen, Martial Hebert, and Rahul Sukthankar. Trajectons: Action recognition through the motion analysis of tracked features. In *2009 IEEE 12th international conference on computer vision workshops, ICCV workshops*, pages 514–521. IEEE, 2009.

[36] Jose M Chaquet, Enrique J Carmona, and Antonio Fernández-Caballero. A survey of video datasets for human action and activity recognition. *Computer Vision and Image Understanding*, 117(6):633–659, 2013.

[37] Temitayo Olugbade, Marta Bieńkiewicz, Giulia Barbareschi, Vincenzo D'amato, Luca Oneto, Antonio Camurri, Catherine Holloway, Mårten Björkman, Peter Keller, Martin Clayton, et al. Human movement datasets: An interdisciplinary scoping review. *ACM Computing Surveys*, 55(6):1–29, 2022.

[38] Christian Schuldt, Ivan Laptev, and Barbara Caputo. Recognizing human actions: a local svm approach. In *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, volume 3, pages 32–36. IEEE, 2004.

[39] Moshe Blank, Lena Gorelick, Eli Shechtman, Michal Irani, and Ronen Basri. Actions as space-time shapes. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, volume 2, pages 1395–1402. IEEE, 2005.

[40] Jingen Liu, Jiebo Luo, and Mubarak Shah. Recognizing realistic actions from videos "in the wild". In *2009 IEEE conference on computer vision and pattern recognition*, pages 1996–2003. IEEE, 2009.

[41] Kishore K Reddy and Mubarak Shah. Recognizing 50 human action categories of web videos. *Machine vision and applications*, 24(5):971–981, 2013.

[42] Khurram Soomro, Amir Roshan Zamir, and Mubarak Shah. Ucf101: A dataset of 101 human actions classes from videos in the wild. *arXiv preprint arXiv:1212.0402*, 2012.

[43] Will Kay, Joao Carreira, Karen Simonyan, Brian Zhang, Chloe Hillier, Sudheendra Vijayanarasimhan, Fabio Viola, Tim Green, Trevor Back, Paul Natsev, et al. The kinetics human action video dataset. *arXiv preprint arXiv:1705.06950*, 2017.

[44] Joao Carreira, Eric Noland, Andras Banki-Horvath, Chloe Hillier, and Andrew Zisserman. A short note about kinetics-600. *arXiv preprint arXiv:1808.01340*, 2018.

[45] Joao Carreira, Eric Noland, Chloe Hillier, and Andrew Zisserman. A short note on the kinetics-700 human action dataset. *arXiv preprint arXiv:1907.06987*, 2019.

[46] Hildegard Kuehne, Hueihan Jhuang, Estíbaliz Garrote, Tomaso Poggio, and Thomas Serre. Hmdb: a large video database for human motion recognition. In *2011 International conference on computer vision*, pages 2556–2563. IEEE, 2011.

[47] Marcin Marszalek, Ivan Laptev, and Cordelia Schmid. Actions in context. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2929–2936. IEEE, 2009.

[48] Serena Yeung, Olga Russakovsky, Ning Jin, Mykhaylo Andriluka, Greg Mori, and Li Fei-Fei. Every moment counts: Dense detailed labeling of actions in complex videos. *International Journal of Computer Vision*, 126:375–389, 2018.

[49] Juan Carlos Niebles, Chih-Wei Chen, and Li Fei-Fei. Modeling temporal structure of decomposable motion segments for activity classification. In *Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part II 11*, pages 392–405. Springer, 2010.

[50] Ming Cheng, Kunjing Cai, and Ming Li. Rwf-2000: an open large scale video database for violence detection. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 4183–4190. IEEE, 2021.

[51] Waqas Sultani, Chen Chen, and Mubarak Shah. Real-world anomaly detection in surveillance videos. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6479–6488, 2018.

[52] Marek Kulbacki, Jakub Segen, Kamil Wereszczyński, and Adam Gudyś. Vmass: massive dataset of multi-camera video for learning, classification and recognition of human actions. In *Intelligent Information and Database Systems: 6th Asian Conference, ACIIDS 2014, Bangkok, Thailand, April 7-9, 2014, Proceedings, Part II 6*, pages 565–574. Springer, 2014.

[53] Jun Liu, Amir Shahroudy, Mauricio Perez, Gang Wang, Ling-Yu Duan, and Alex C Kot. Ntu rgb+ d 120: A large-scale benchmark for 3d human activity understanding. *IEEE transactions on pattern analysis and machine intelligence*, 42(10):2684–2701, 2019.

[54] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.

[55] Warren S McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5:115–133, 1943.

[56] Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386, 1958.

[57] Kunihiko Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics*, 36(4):193–202, 1980.

[58] Michael I Jordan. Serial order: A parallel distributed processing approach. In *Advances in psychology*, volume 121, pages 471–495. Elsevier, 1986.

[59] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[60] Geoffrey E Hinton, Simon Osindero, and Yee-Whye Teh. A fast learning algorithm for deep belief nets. *Neural computation*, 18(7):1527–1554, 2006.

[61] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814, 2010.

[62] Erfan Azarkhish, Davide Rossi, Igor Loi, and Luca Benini. Neurostream: Scalable and energy efficient deep learning with smart memory cubes. *IEEE Transactions on Parallel and Distributed Systems*, 29(2):420–434, 2017.

[63] H Brendan McMahan, Eider Moore, Daniel Ramage, and Blaise Agüera y Arcas. Federated learning of deep networks using model averaging. *arXiv preprint arXiv:1602.05629*, 2:2, 2016.

[64] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017.

[65] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[66] Henry J Kelley. Gradient theory of optimal flight paths. *Ars Journal*, 30(10):947–954, 1960.

[67] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *nature*, 323(6088):533–536, 1986.

[68] Herbert Robbins and Sutton Monro. A stochastic approximation method. *The annals of mathematical statistics*, pages 400–407, 1951.

[69] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[70] John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of machine learning research*, 12(7), 2011.

[71] Tijmen Tieleman. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural networks for machine learning*, 4(2):26, 2012.

[72] Kunihiko Fukushima. Neocognitron: A hierarchical neural network capable of visual pattern recognition. *Neural networks*, 1(2):119–130, 1988.

[73] Wenchao Jiang and Zhaozheng Yin. Human activity recognition using wearable sensors by deep convolutional neural networks. In *Proceedings of the 23rd ACM international conference on Multimedia*, pages 1307–1310, 2015.

[74] Jeffrey L Elman. Finding structure in time. *Cognitive science*, 14(2):179–211, 1990.

[75] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[76] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 31, 2017.

[77] Stevo Bozinovski. Reminder of the first paper on transfer learning in neural networks, 1976. *Informatica*, 44(3), 2020.

[78] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.

[79] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[80] Andrew Jaegle, Felix Gimeno, Andy Brock, Oriol Vinyals, Andrew Zisserman, and Joao Carreira. Perceiver: General perception with iterative attention. In *International conference on machine learning*, pages 4651–4664. PMLR, 2021.

[81] Midori Kitagawa and Brian Windsor. *MoCap for artists: workflow and techniques for motion capture*. Elsevier/Focal Press, Amsterdam ; Boston, 2008. OCLC: ocn190620556.

[82] Alberto Menache. *Understanding motion capture for computer animation*. Morgan Kaufmann, Burlington, MA, 2nd edition, 2011.

[83] Przemysław Skurowski and Magdalena Pawlyta. On the noise complexity in an optical motion capture facility. *Sensors*, 19(20):4435, 2019.

[84] David W Allan. Statistics of atomic frequency standards. *Proceedings of the IEEE*, 54(2):221–230, 1966.

[85] Adam Świtoński, Magdalena Stawarz, Magdalena Boczarska-Jedynak, Aleksander Sieroń, Andrzej Polański, and Konrad Wojciechowski. The effectiveness of applied treatment in parkinson disease based on feature selection of motion activities. *Przegląd Elektrotechniczny*, 88(12B):103–106, 2012.

[86] Adam Świtoński, Magdalena Stawarz, Aleksander Sieroń, Andrzej Polański, and Konrad Wojciechowski. Skuteczność leczenia w chorobie parkinsona na bazie selekcji charakterystycznych cech chodu.

[87] Andrzej W Przybyszewski, Magdalena Boczarska, Stanisław Kwiek, and Konrad Wojciechowski. Rough set based classifications of parkinson's patients gaits. In *Intelligent Information and Database Systems: 6th Asian Conference, ACIIDS 2014, Bangkok, Thailand, April 7-9, 2014, Proceedings, Part II 6*, pages 525–534. Springer, 2014.

[88] Magdalena Stawarz, Andrzej Polański, Stanisław Kwiek, Magdalena Boczarska-Jedynak, Łukasz Janik, Andrzej Przybyszewski, and Konrad Wojciechowski. A system for analysis of tremor in patients with parkinson's disease based on motion capture technique. In *Computer Vision and Graphics: International Conference, ICCVG 2012, Warsaw, Poland, September 24-26, 2012. Proceedings*, pages 618–625. Springer, 2012.

[89] Monika Błaszczyszyn, Agnieszka Szczęsna, Magdalena Pawlyta, Maciej Marszałek, and Dariusz Karczmit. Kinematic analysis of mae-geri kicks in beginner and advanced kyokushin karate athletes. *International journal of environmental research and public health*, 16(17):3155, 2019.

[90] Jacek Wąsik and Tomasz Góra. Impact of target selection on front kick kinematics in taekwondo–pilot study. *Physical Activity Review*, 4:57–61, 2016.

[91] Jacek Wąsik, Dorota Ortenburger, Tomasz Góra, and Dariusz Mosler. The influence of effective distance on the impact of a punch–preliminary analysis. *Physical Activity Review*, 6:81–86, 2018.

[92] Jacek Wąsik, Dorota Ortenburger, and Tomasz Góra. Physiotherapeutic applications of biomechanical opposing indicators–based on measurements of taekwon-do athlets. *Physical education, sport and health culture in modern society*, (2 (38)):212–215, 2017.

[93] Jacek Wąsik, Tomasz Góra, Dorota Ortenburger, and Gongbing Shan. Kinematic quantification of straight-punch techniques using the preferred and non-preferred fist in taekwon-do. *Biomedical Human Kinetics*, 11(1):115–120, 2019.

[94] Agnieszka Szczęsna, Monika Błaszczyszyn, and Magdalena Pawlyta. Optical motion capture dataset of selected techniques in beginner and advanced kyokushin karate athletes. *Scientific Data*, 8(1):13, 2021.

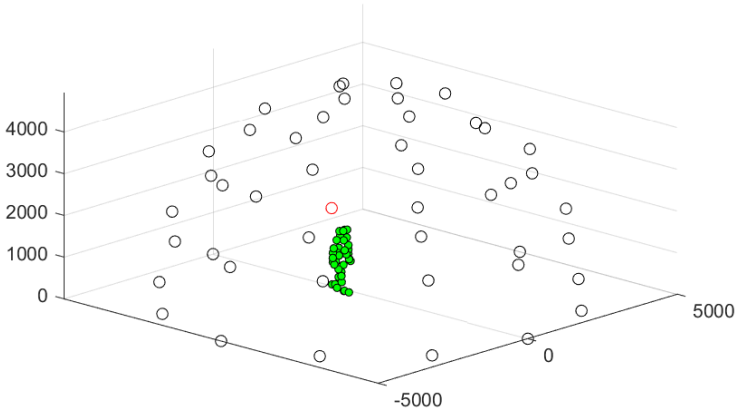
[95] Przemysław Skurowski and Magdalena Pawlyta. Detection and classification of artifact distortions in optical motion capture sequences. *Sensors*, 22(11):4076, 2022.

[96] Przemysław Skurowski and Magdalena Pawlyta. Tree based regression methods for gap reconstruction of motion capture sequences. *Biomedical Signal Processing and Control*, 88:105641, 2024.

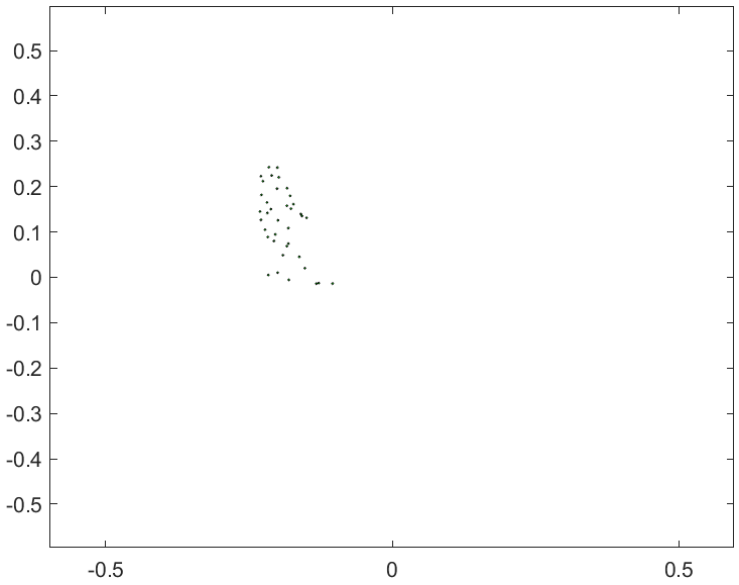
[97] Jan Gorodkin. Comparing two k-category assignments by a k-category correlation coefficient. *Computational biology and chemistry*, 28(5-6):367–374, 2004.

[98] Jonas Močkus. On bayesian methods for seeking the extremum. In *Optimization techniques IFIP technical conference: Novosibirsk, July 1–7, 1974*, pages 400–404. Springer, 1975.

A Wybrane rzuty perspektywiczne

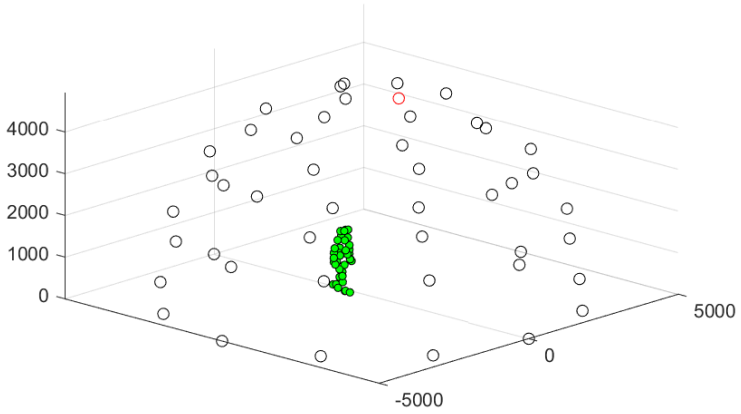


(a)

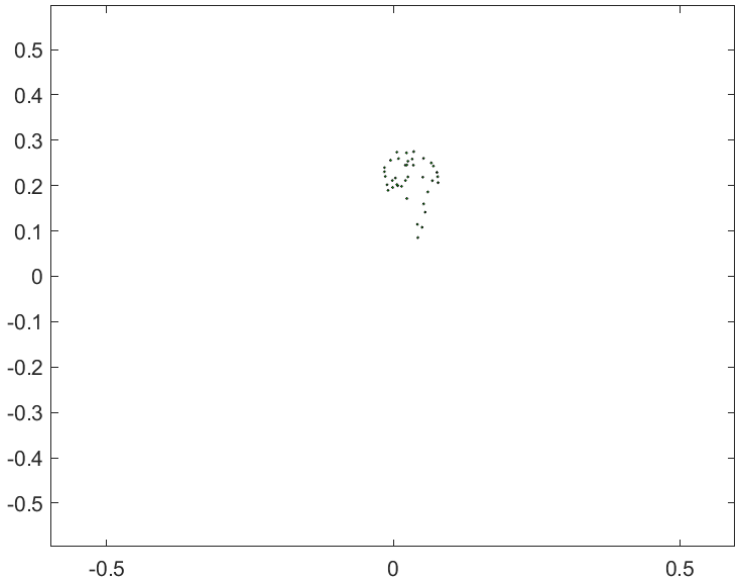


(b)

Rysunek 143: Wybrany rzut perspektywiczny (b) dla zaznaczonej pozycji wirtualnej kamery (a)

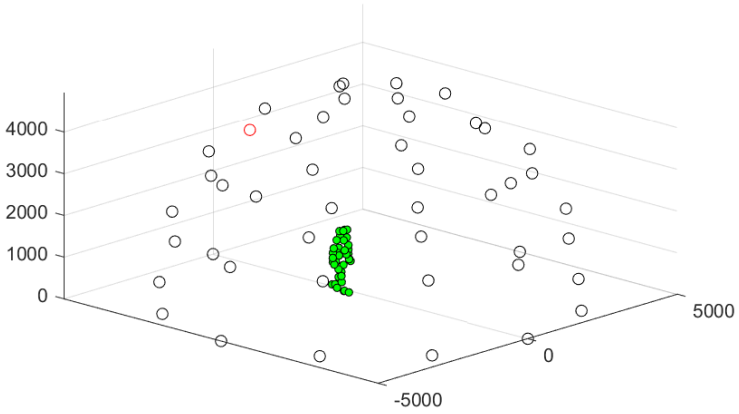


(a)

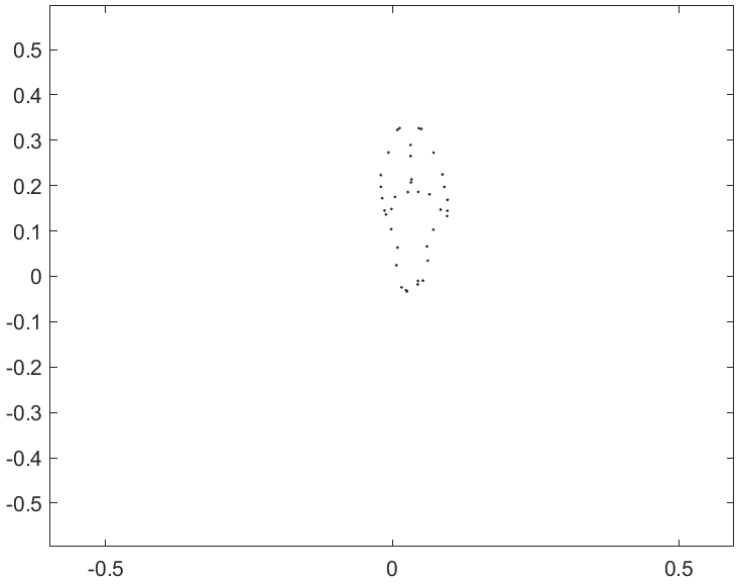


(b)

Rysunek 144: Wybrany rzut perspektywiczny (b) dla zaznaczonej pozycji wirtualnej kamery (a)



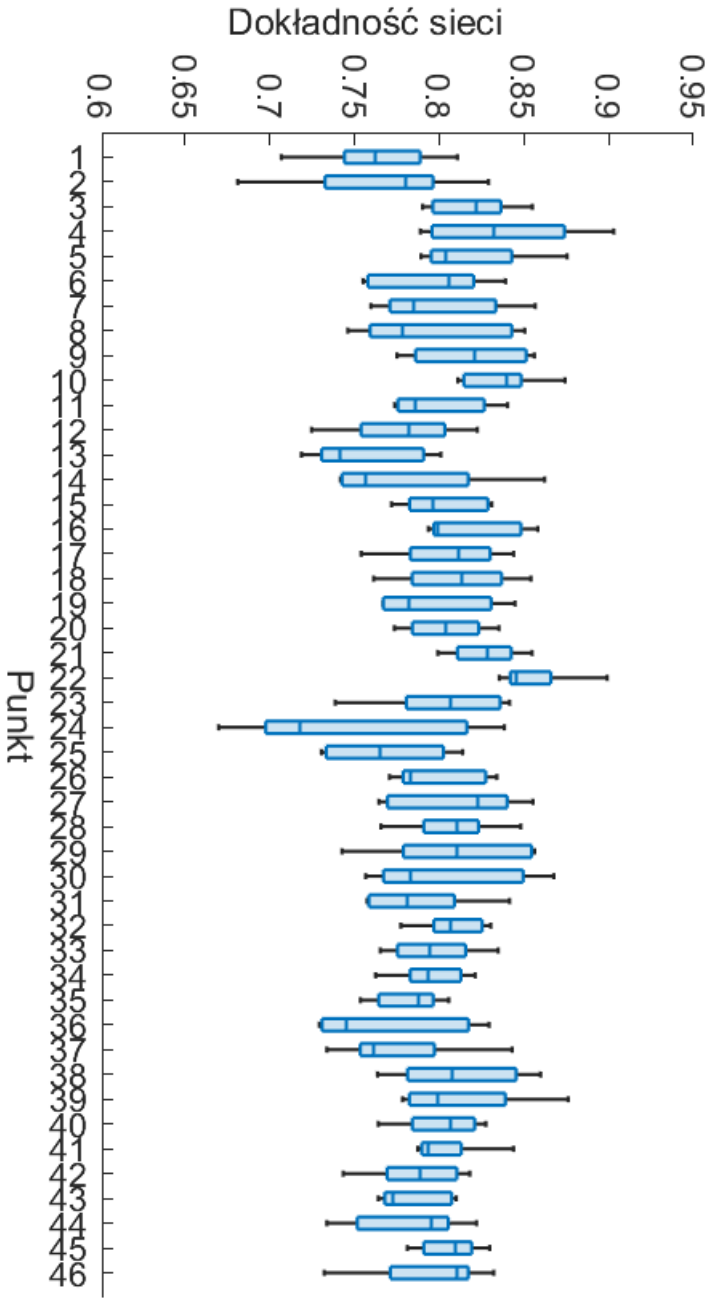
(a)



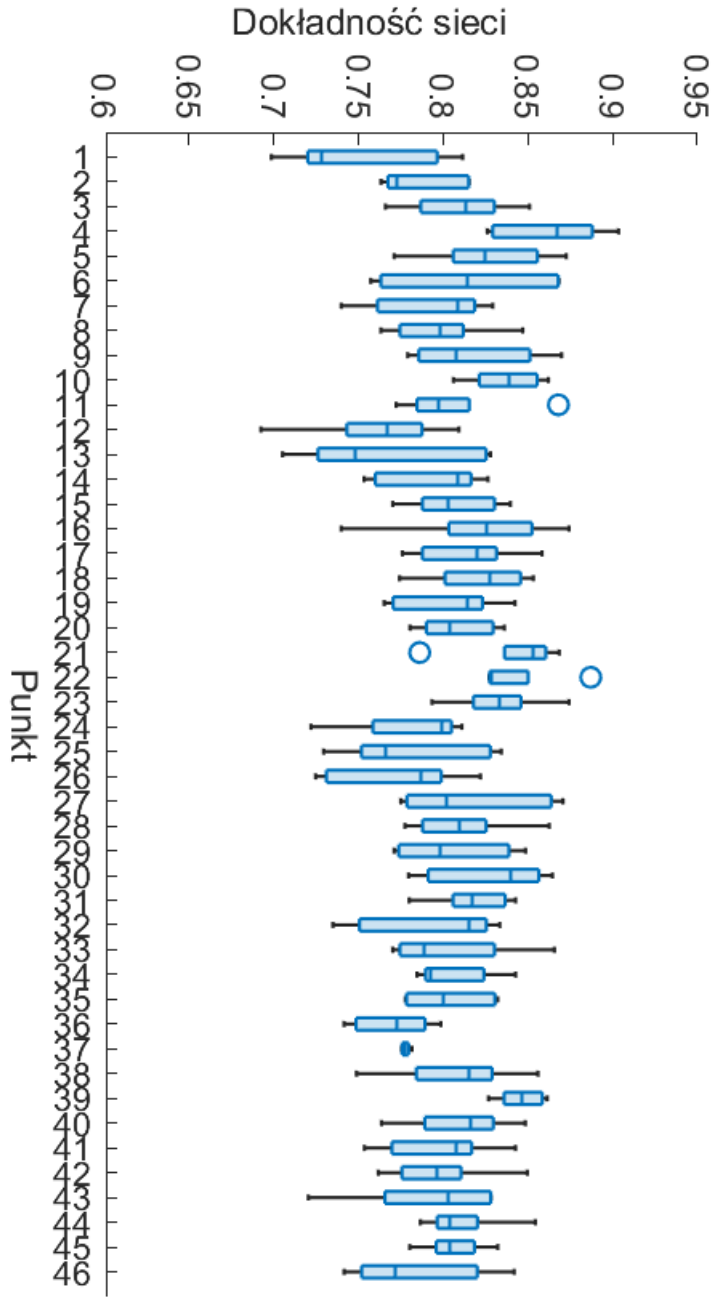
(b)

Rysunek 145: Wybrany rzut perspektywiczny (b) dla zaznaczonej pozycji wirtualnej kamery (a)

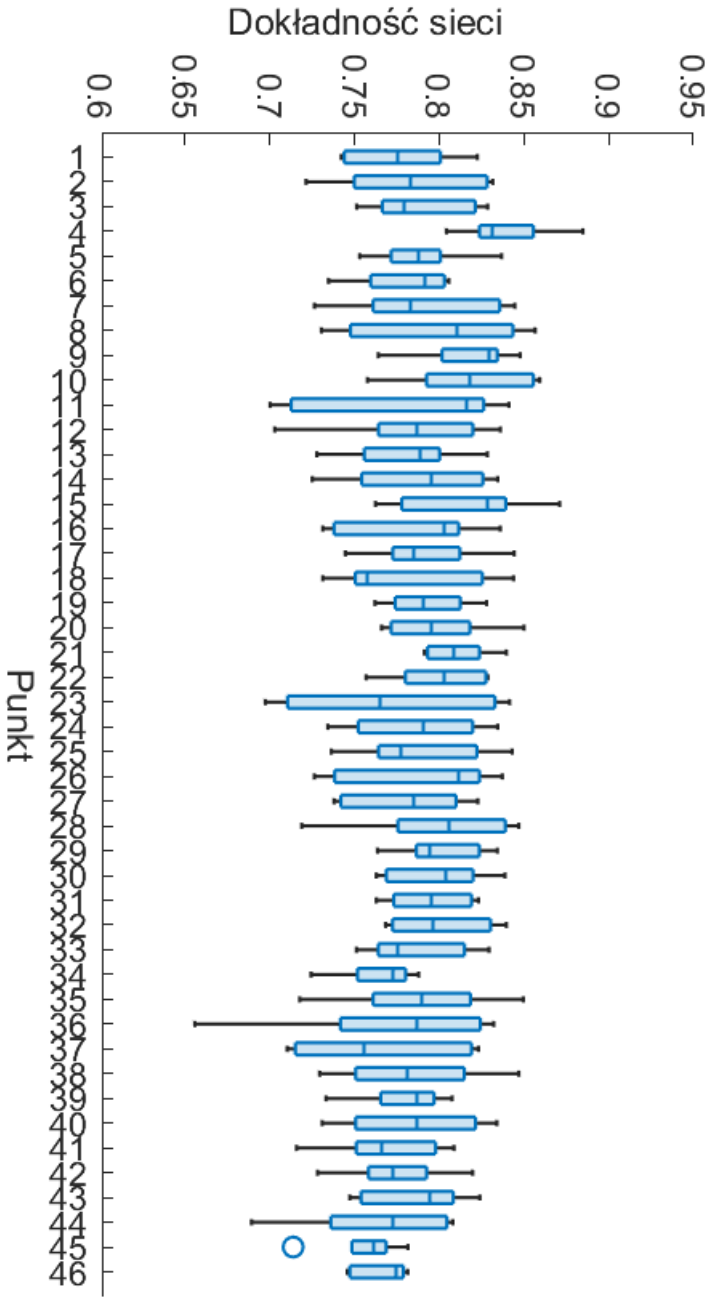
B Wykresy pudełkowe dla sieci LSTM



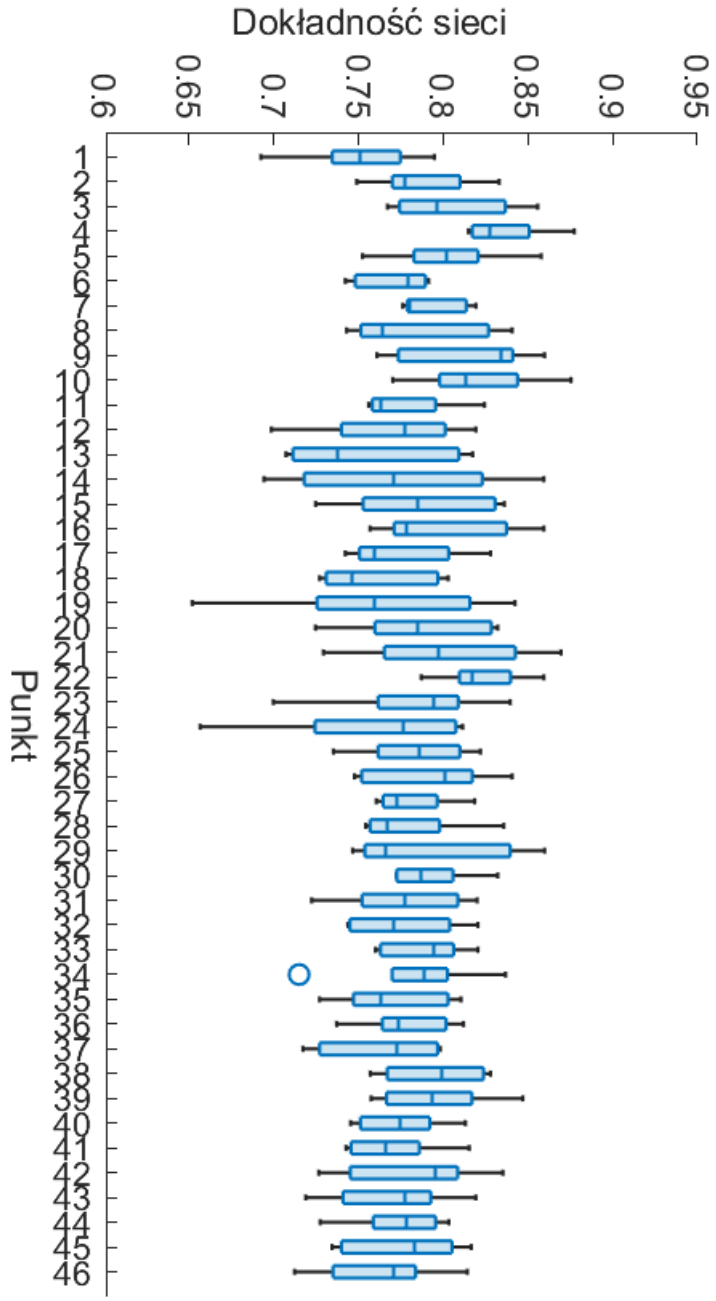
Rysunek 146: Wykres pudełkowy dokładności sieci dla wszystkich położeń wirtualnej kamery, dla 38 markerów wektorze wejściowym



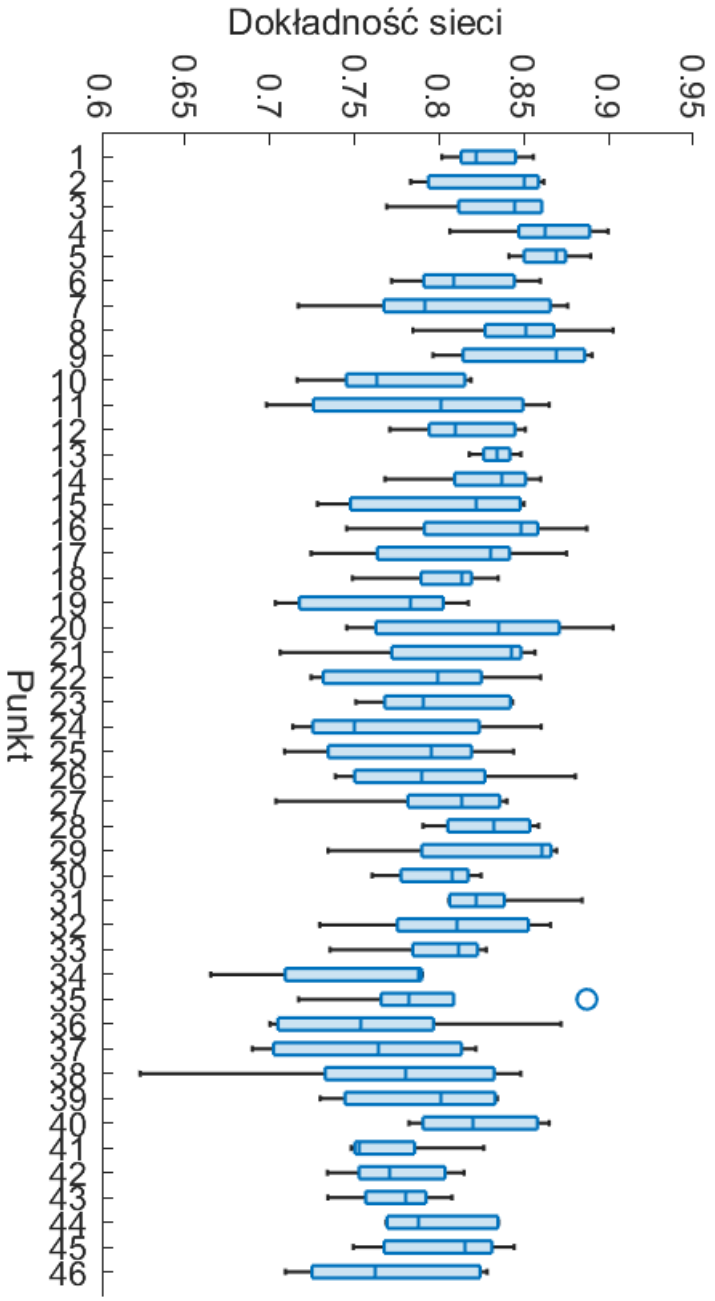
Rysunek 147: Wykres pudełkowy dokładności sieci dla wszystkich położeń wirtualnej kamery, dla 28 markerów wektorze wejściowym



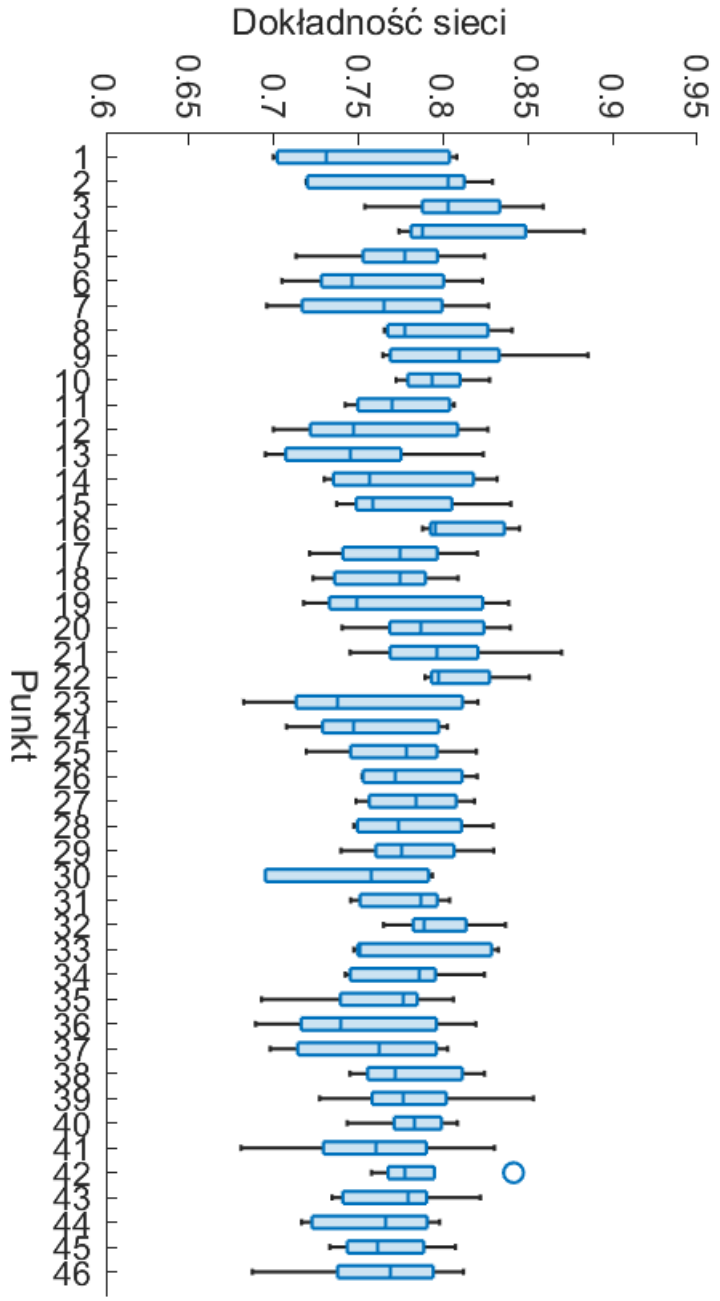
Rysunek 148: Wykres pudełkowy dokładności sieci dla wszystkich położeń wirtualnej kamery, dla 22 markerów wektorze wejściowym



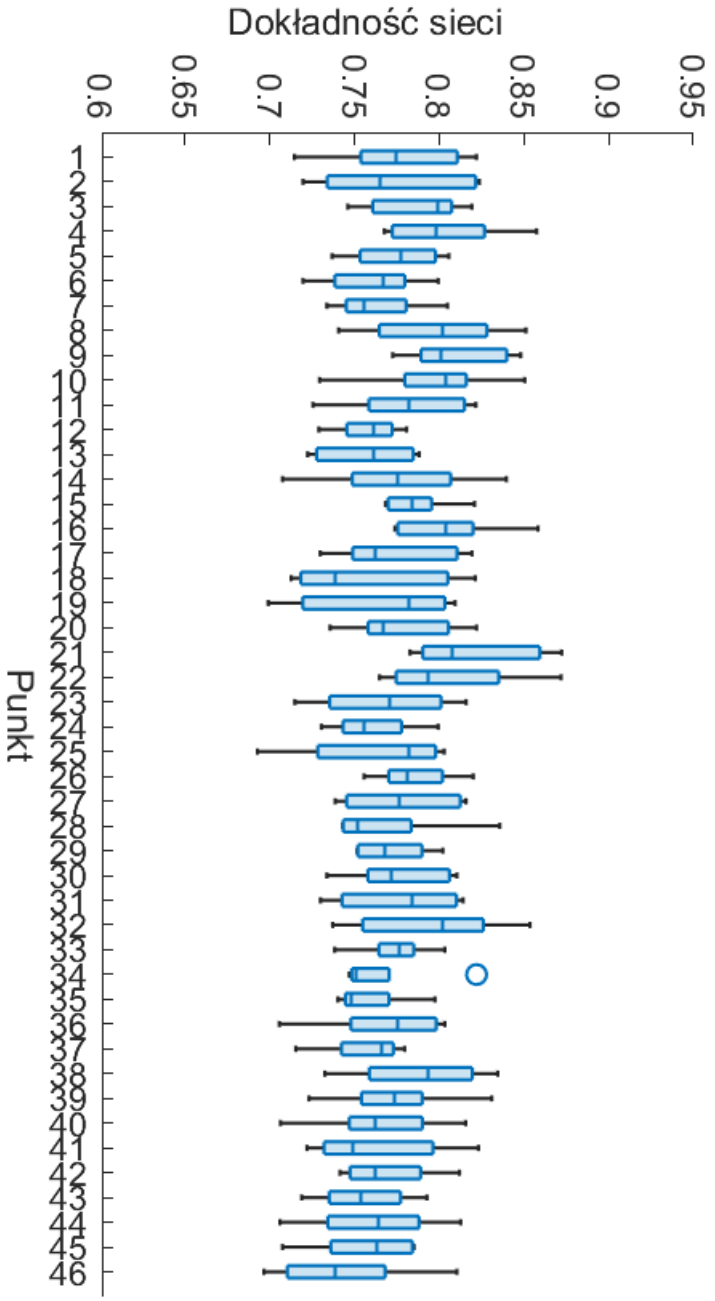
Rysunek 149: Wykres pudełkowy dokładności sieci dla wszystkich położeń wirtualnej kamery, dla 16 markerów wektorze wejściowym



Rysunek 150: Wykres pudełkowy dokładności sieci dla wszystkich położeń wirtualnej kamery, dla 13 markerów wektorze wejściowym

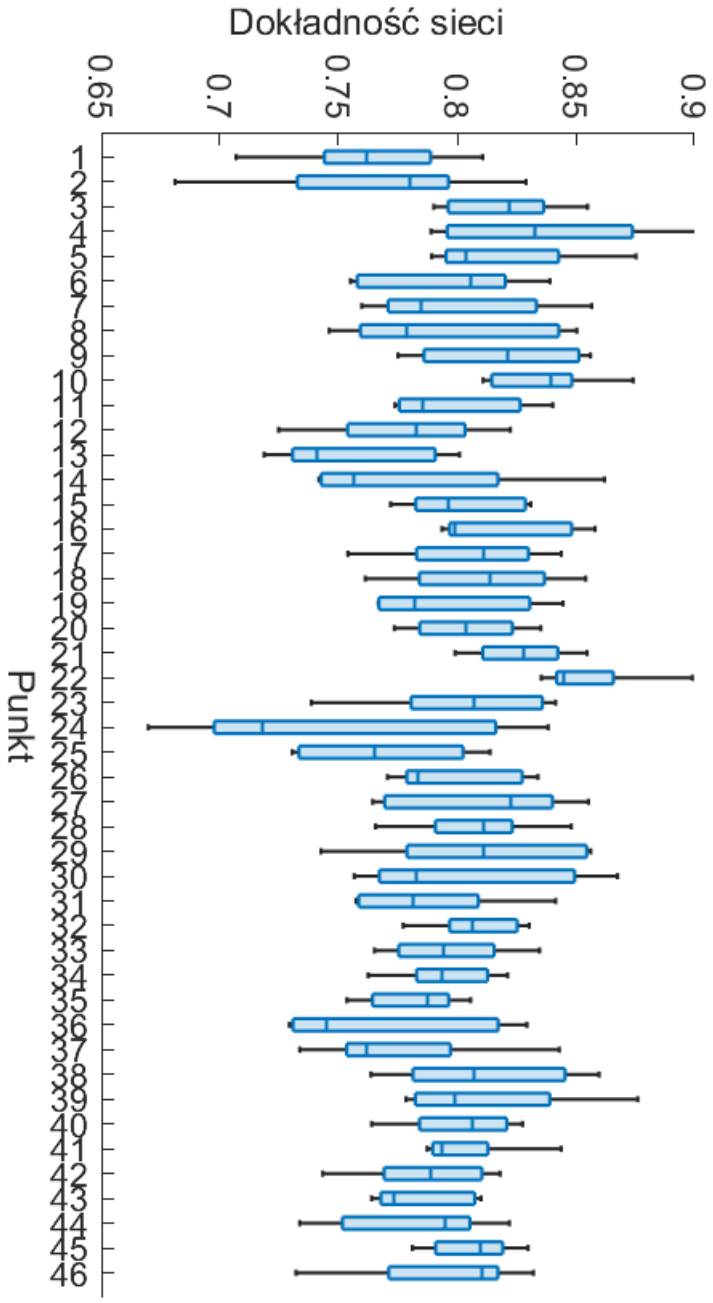


Rysunek 151: Wykres pudełkowy dokładności sieci dla wszystkich położeń wirtualnej kamery, dla 9 markerów wektorze wejściowym

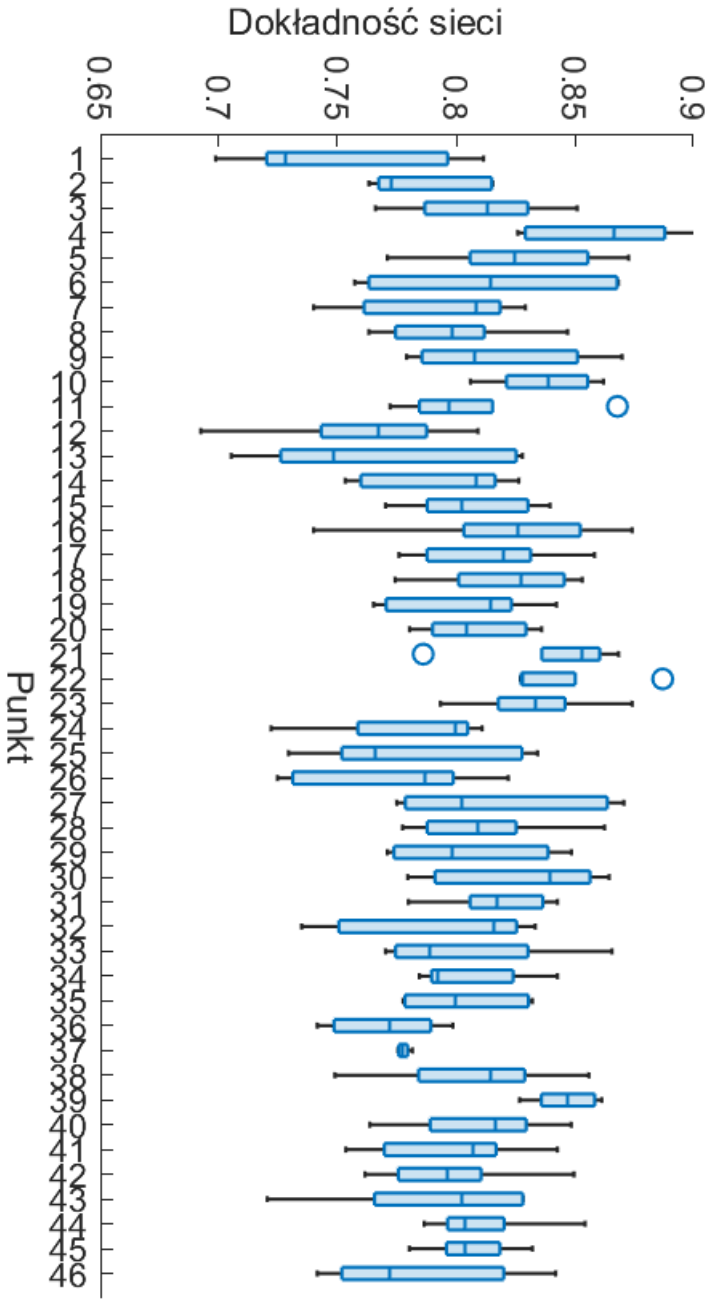


Rysunek 152: Wykres pudełkowy dokładności sieci dla wszystkich położeń wirtualnej kamery, dla 6 markerów wektorze wejściowym

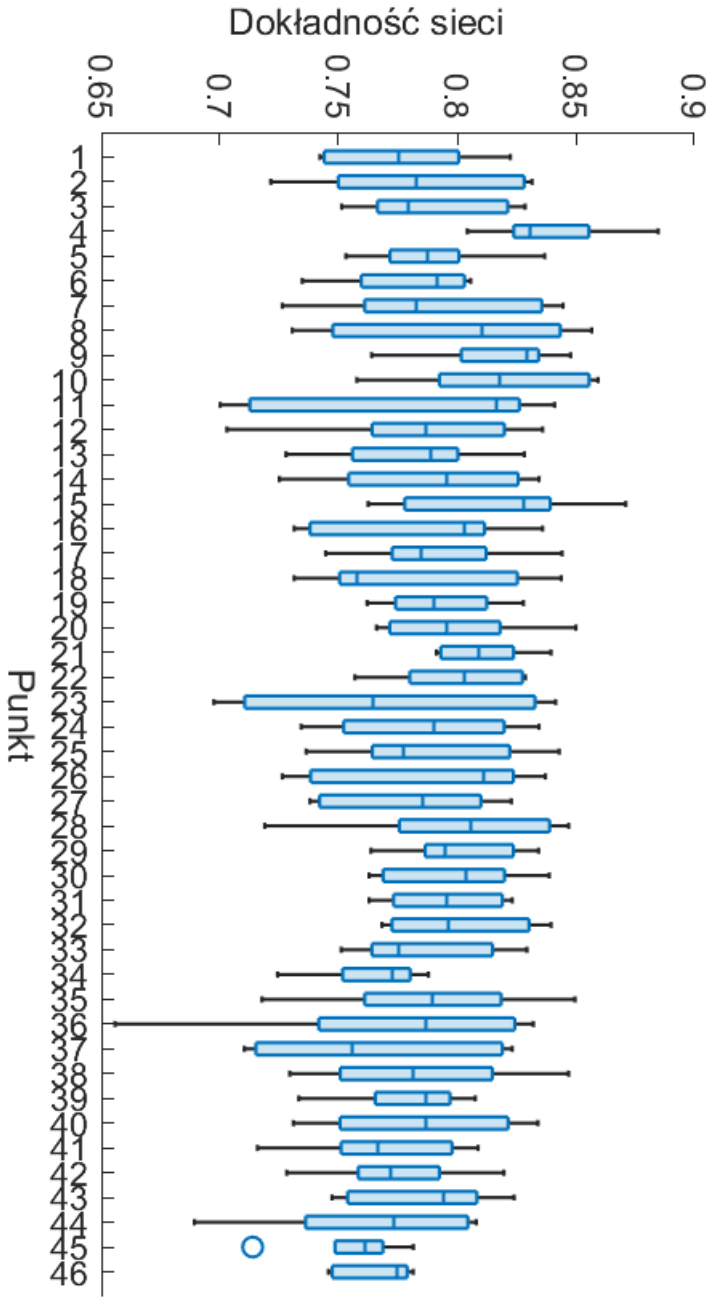
C Wykresy pudełkowe dla sieci CNN



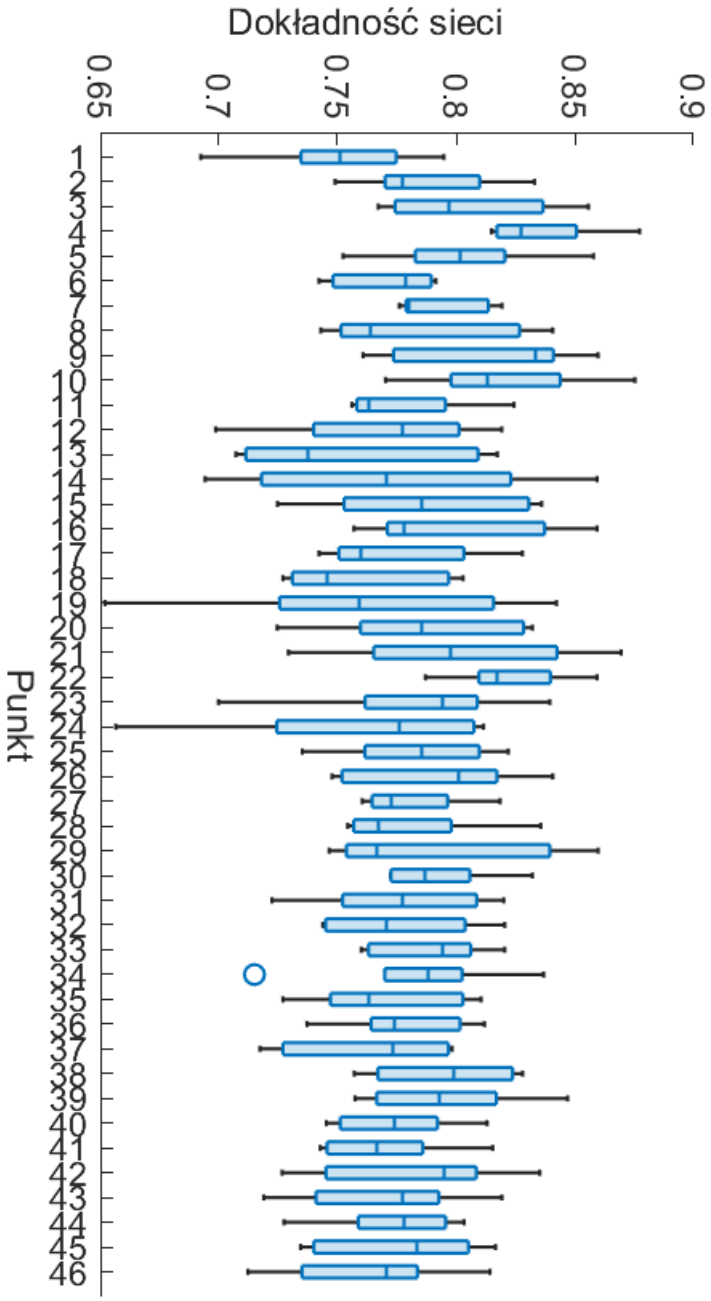
Rysunek 153: Wykres pudełkowy dokładności sieci dla wszystkich położeń wirtualnej kamery, dla 38 markerów wektorze wejściowym



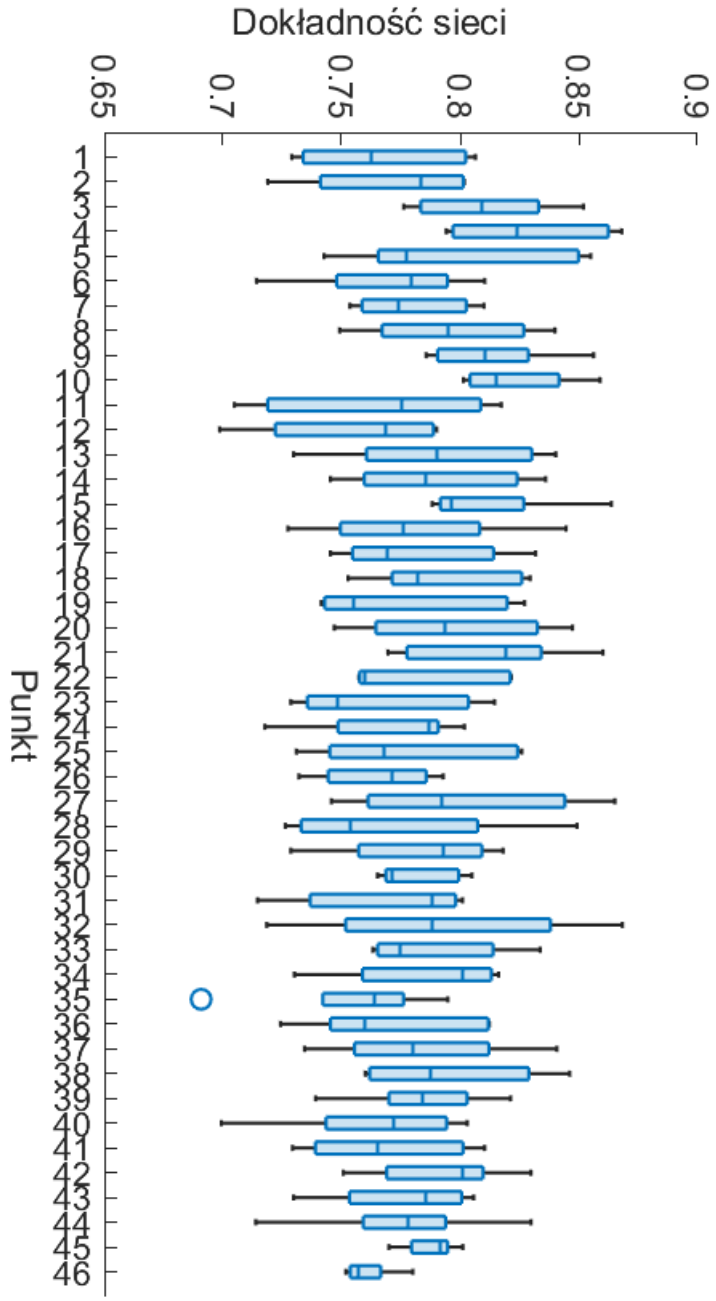
Rysunek 154: Wykres pudełkowy dokładności sieci dla wszystkich położeń wirtualnej kamery, dla 28 markerów wektorze wejściowym



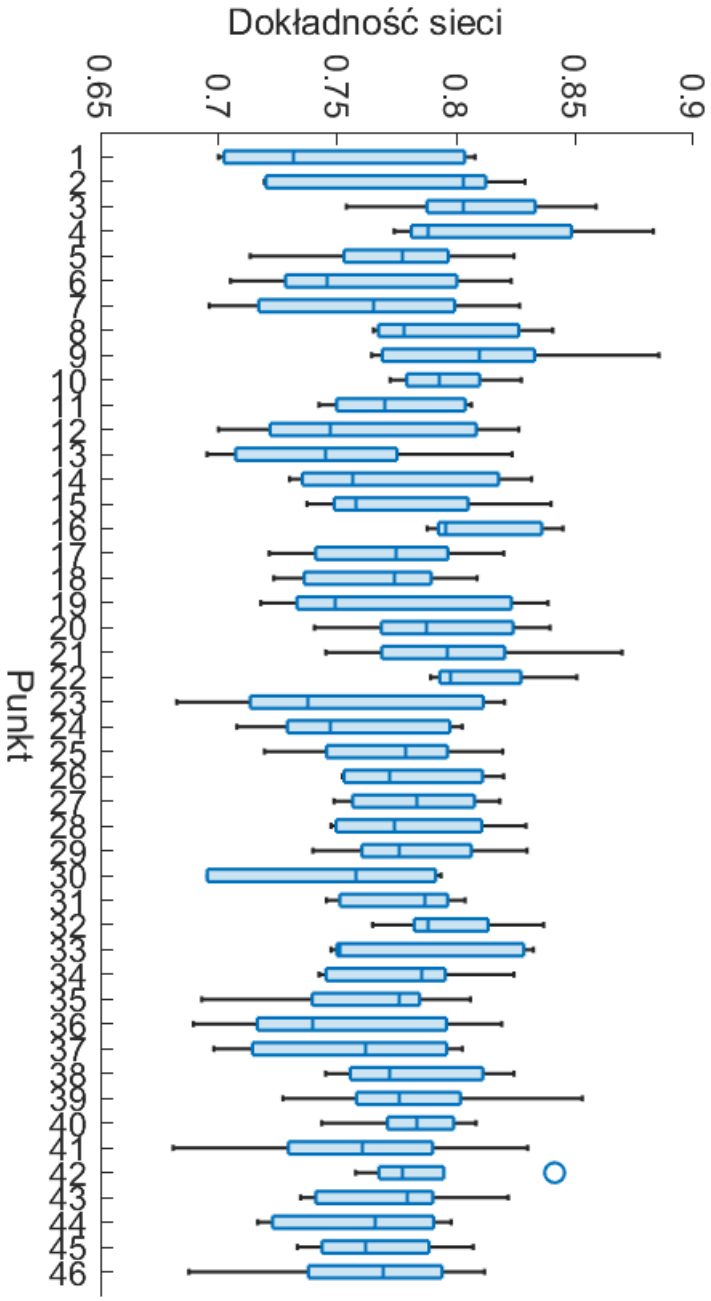
Rysunek 155: Wykres pudełkowy dokładności sieci dla wszystkich położeń wirtualnej kamery, dla 22 markerów wektorze wejściowym



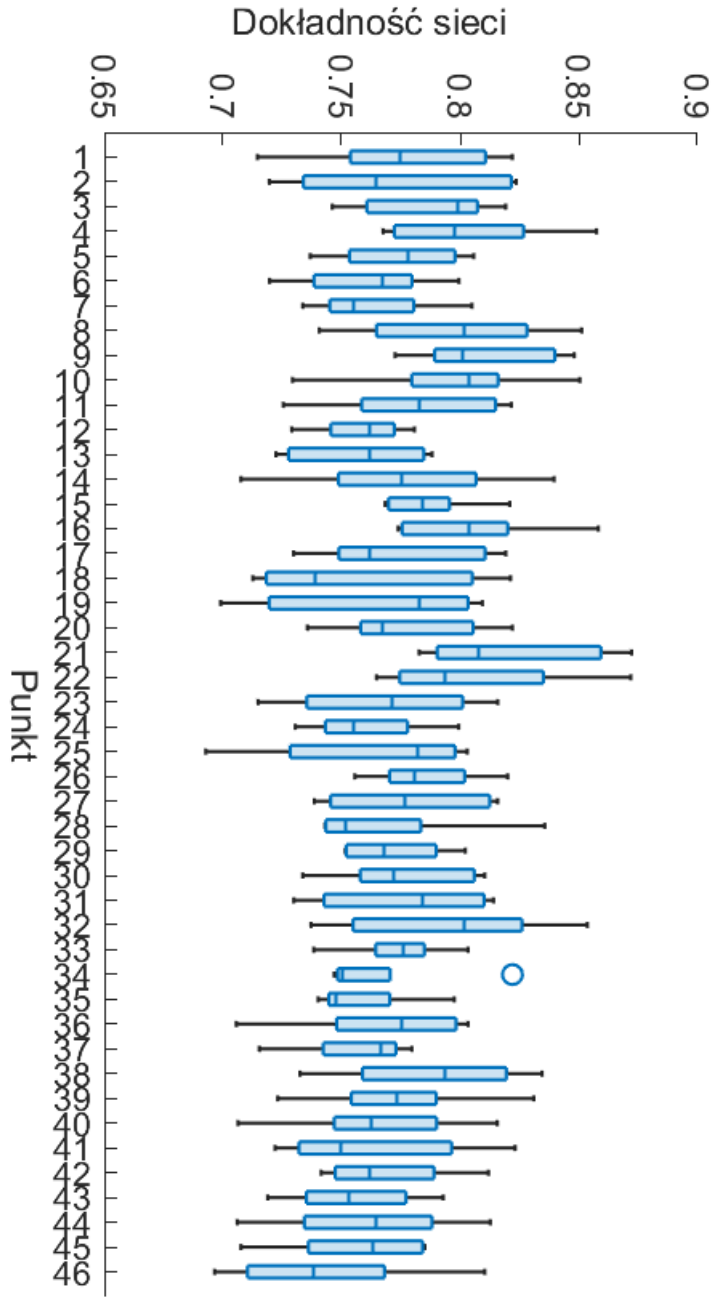
Rysunek 156: Wykres pudełkowy dokładności sieci dla wszystkich położeń wirtualnej kamery, dla 16 markerów wektorze wejściowym



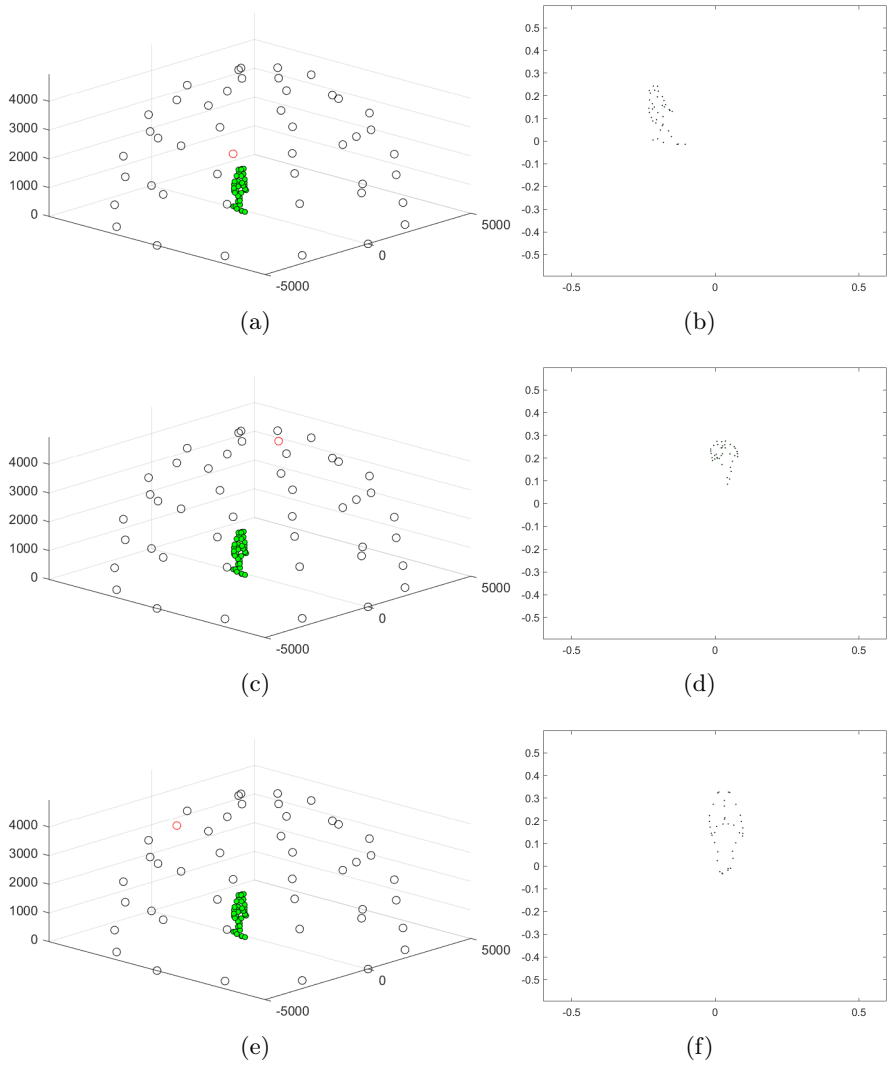
Rysunek 157: Wykres pudełkowy dokładności sieci dla wszystkich położeń wirtualnej kamery, dla 13 markerów wektorze wejściowym



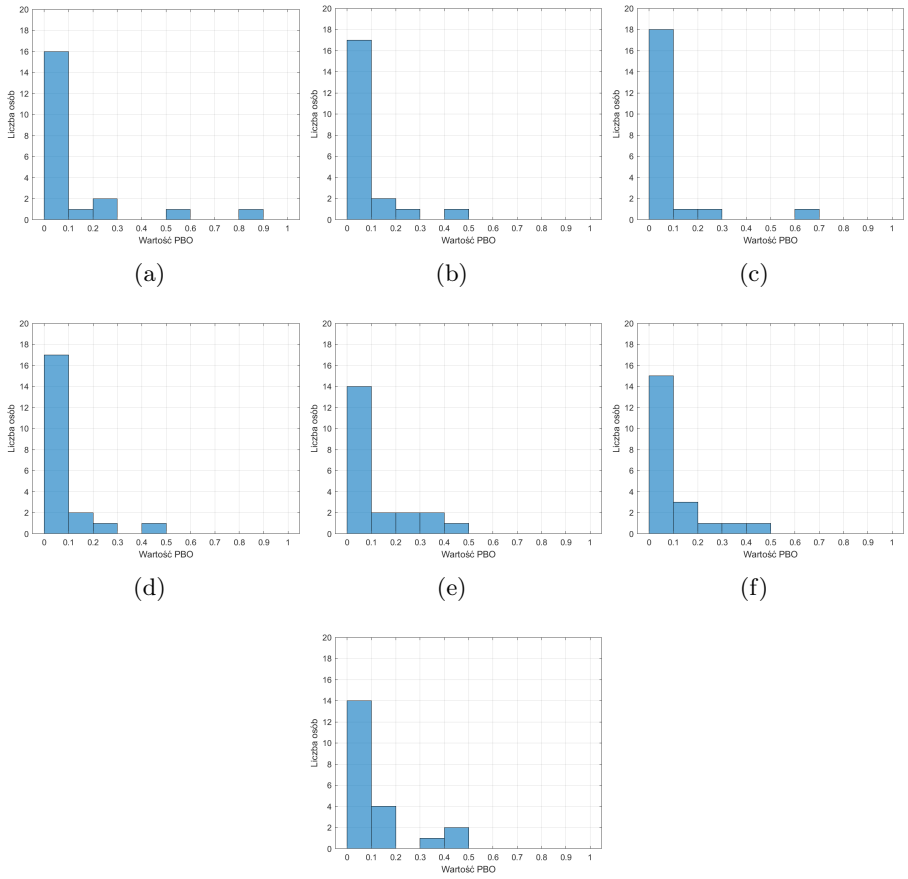
Rysunek 158: Wykres pudełkowy dokładności sieci dla wszystkich położeń wirtualnej kamery, dla 9 markerów wektorze wejściowym



Rysunek 159: Wykres pudełkowy dokładności sieci dla wszystkich położeń wirtualnej kamery, dla 6 markerów wektorze wejściowym



Rysunek 27: Wybrane rzuty perspektywiczne (b,d,f) dla trzech różnych pozycji wirtualnej kamery (a,c,e).

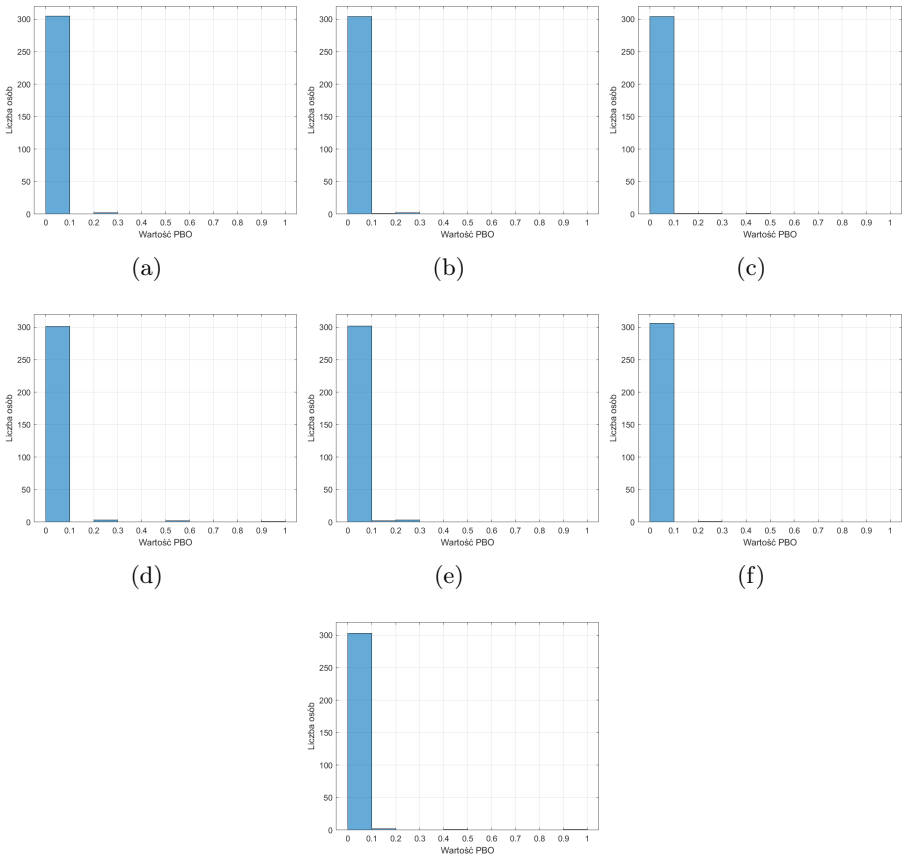


(a) (b) (c)

(d) (e) (f)

(g)

Rysunek 40: Rozkład PBO dla akcji Obrót dla różnej liczby znaczników w wektorze wejściowym: (a) 38 znaczników, (b) 28 znaczników, (c) 22 znaczniki, (d) 16 znaczników, (e) 13 znaczników, (f) 9 znaczników, (g) 6 znaczników.

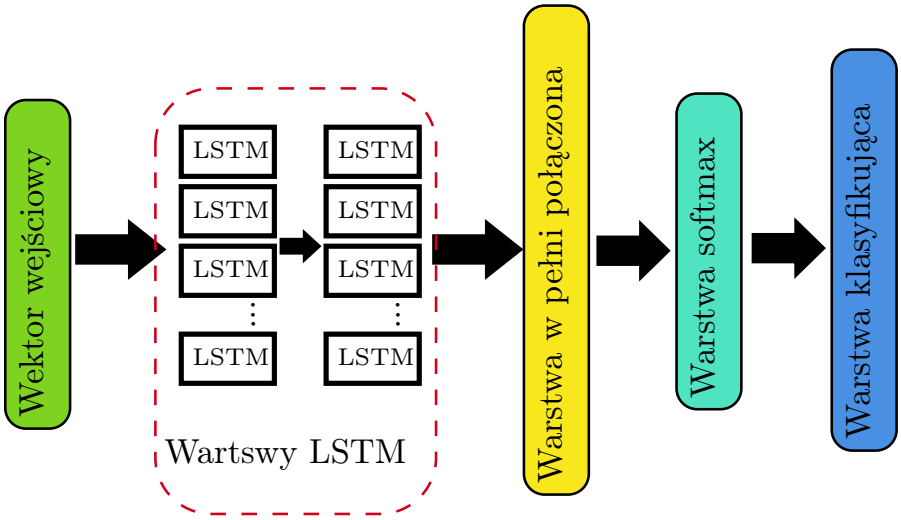


(a) (b) (c)

(d) (e) (f)

(g)

Rysunek 41: Rozkład PBO dla akcji Schylanie się dla różnej liczby znaczników w wektorze wejściowym: (a) 38 znaczników, (b) 28 znaczników, (c) 22 znaczniki, (d) 16 znaczników, (e) 13 znaczników, (f) 9 znaczników, (g) 6 znaczników.



Rysunek 59: Schemat zaprojektowanej sieci LSTM