



POLSKA AKADEMIA NAUK
INSTYTUT BIOCYBERNETYKI I INŻYNIERII BIOMEDYCZNEJ

Praca doktorska

Proces gojenia ściegna Achillesa oceniany przez fuzję danych z
wykorzystaniem głębkich sieci neuronowych

Autor: mgr inż. Norbert Kapiński
Kierujący pracą: dr hab. inż. Antoni Grzanka
Promotor pomocniczy: dr Jakub Zieliński

Warszawa, wrzesień 2018

Streszczenie

The abstract will go here....

W tym miejscu można umieścić abstrakt pracy. W przeciwnym wypadku należy usunąć/zakomentować niniejszy fragment kodu.

Spis treści

1 Wstęp	1
2 Cel i przebieg pracy	5
3 Monitorowanie procesu gojenia ścięgna Achillesa	7
3.1 Ściegno Achillesa	7
3.1.1 Anatomia	8
3.1.2 Biomechanika	10
3.1.3 Urazy i czynniki im sprzyjające	12
3.1.4 Fazy gojenia, leczenie i rehabilitacja	15
3.2 Zastosowanie rezonansu magnetycznego	17
3.3 Zastosowanie ultrasonografii	23
3.4 Zastosowanie badań biomechanicznych	28
4 Konwolucyjne sieci neuronowe	32
4.1 Zarys historyczny	35
4.2 Szkolenie głębokich sieci neuronowych	38
4.2.1 Problem nadmiernego dopasowania	43
4.2.2 Problem redukcji wymiarowości	46
4.3 Przykłady współczesnych topologii	50
4.3.1 AlexNet	52

4.3.2	GoogLeNet	55
4.3.3	ResNet	57
4.3.4	Złożenia	60
4.4	Zastosowania w medycynie	62
5	Nowa metoda oceny procesu gojenia ścięgna Achillesa	66
5.1	Metodyka	67
5.1.1	Zbiór danych	69
5.1.2	Wzorzec odniesienia	72
5.2	Eksperymenty	73
5.2.1	Rozróżnienie ścięgna zdrowego i po zerwaniu	74
5.2.2	Obliczanie i redukcja wymiarowości wektora cech	76
5.2.3	Wybór sekwencji rezonansu magnetycznego	78
5.2.4	Optymalna selekcja i konkatenacja cech	84
5.3	Ocena procesu gojenia	87
5.3.1	Macierze błędu	87
5.3.2	Krzywe gojenia	92
6	Testy nowej metody	98
6.1	Porównanie z metodą opartą o paradymat end-to-end	98
6.2	Porównanie z metodą opartą o dane z ultrasonografii	101
6.3	Porównanie z metodą opartą o badania biomechaniczne	105
7	Podsumowanie	111
Bibliografia		127
A AchillesDL: System komputerowego wspomagania oceny gojenia ścięgien i więzadeł		128

Spis rysunków

1.1	Globalny udział w PKB rozwiązań opartych o SI aplikowanych na rynku Med-Tech.	2
1.2	Podział przedstawiający różne rodzaje współczesnych architektur głębokich sieci neuronowych.	3
3.1	Lokalizacja mięśnia trójdźwiedzkiego łydki wraz ze ścięgnem Achillesa.	7
3.2	Schemat hierarchicznej budowy ścięgna Achillesa.	9
3.3	Schemat generowania skurcza mięśni.	11
3.4	Schemat modelu typu Hilla stosowany w oprogramowaniu OpenSim.	11
3.5	Wykres przedstawiający proporcje występowania urazu ścięgna Achillesa w różnych dyscyplinach sportowych na przykładzie amerykańskiego społeczeństwa (na podst. [101]).	13
3.6	Parametry ośrodków często mierzonych w badaniach USG.	24
4.1	Porównanie schematów przetwarzania danych z wykorzystaniem metod głębokiego uczenia się i innych algorytmów.	35
4.2	Topologia perceptronu.	36
4.3	Topologia perceptronu wielowarstwowego.	37
4.4	Topologia sieci LeNet.	38
4.5	Reprezentacja graficzna oceny krzyżowej.	44

4.6	Dwuwymiarowa przestrzeń cech wraz z zaznaczonymi poziomami warianci i składowymi głównymi.	49
4.7	Błąd top-5 klasyfikacji obiektów w kolejnych latach uzyskiwany przez zwycięzców konkursu ILSVRC	52
4.8	Schemat architektury AlexNet.	53
4.9	Topologia architektury AlexNet z podziałem na dwa akceleratory GPU.	54
4.10	Topologia architektury AlexNet z podziałem na dwa akceleratory GPU.	55
4.11	Topologia architektury GoogleNet	56
4.12	Schemat funkcjonalny pojedynczego bloku w architekturze ResNet.	58
4.13	Topologia architektury ResNet-18.	59
4.14	Statystyki dotyczące publikacji medycznych zawierających słowa kluczowe związane z głębokim uczeniem się.	63
4.15	Porównanie automatycznej klasyfikacji retinopatii cukrzycowej i cukrzycowego obrzęku plamki z oceną panelu ekspertów.	63
4.16	Porównanie automatycznej klasyfikacji 3 chorób skóry z oceną ekspertów dermatologów.	64
5.1	Schemat automatycznej metody oceny procesu gojenia się ścięgna Achillesa.	68
5.2	Wizualizacja danych RM w sekwencji PD dla zdrowego (po lewej) i gojącego się (po prawej) ścięgna Achillesa.	71
5.3	Porównanie czasów treningu architektur na dwóch architekturach GPU: Pascal i Volta.	75
5.4	Porównanie krzywych dla 10 modalności RM reprezentujących średnią H_{PC} z 10 pacjentów.	77
5.5	Znormalizowane histogramy dla 10 sekwencji RM.	79

5.6 Porównanie sekwencji RM ilustrujących proces gojenia się ścięgna Achil- lesa w 12 tygodniu po rekonstrukcji	80
5.7 Proces gojenia się ścięgna Achillea widoczny w obrazach zrekonstru- owanych na podstawie danych z sekwencji T2 * GRE TE_MIN	81
5.8 Analiza wizualna procesu wnioskowania sieci z zaznaczonym obszarem zainteresowania	88
5.9 Macierz średniego błędu absolutnego dla oceny pacjentów testowych.	89
5.10 PP_d	90
5.11 Macierz błędu dla noty jako klasy.	91
5.12 MAE w obrębie klas.	92
5.13 Ocena parametru SCT.	92
5.14 Ocena parametru TT.	93
5.15 Ocena parametru STE.	94
5.16 Ocena parametru TE.	95
5.17 Ocena parametru TU.	96
5.18 Ocena parametru TisE.	97
6.1 Automatyczna segmentacja ROI z wykorzystaniem głębokich sieci neu- ronowych.	101
6.2 Wizualizacja przykładowych danych USG w kolejnych tygodniach po zszyciu ścięgna w przekrojach osiowych i strzałkowych.	103

Spis tabel

4.1	Parametry architektury GoogleNet	56
5.1	Wyniki szkolenia sieci neuronowych z wykorzystaniem kroswalidacji z podziałem na 5 segmentów dla problemu klasyfikacji obrazów zdrowego i chorego ścięgna.	74
5.2	Rezultaty PCA dla 1, 10 i 200 pierwszych czynników głównych.	76
5.3	Korelacja sekwencji PD, T1, T2 and T2* GRE MRI. Wyniki oznaczone pogrubieniem są istotne statystycznie z $p < 0.01$	81
5.4	Porównanie wnioskowania dla różnych modeli regresji wytrenowanych w oparciu o dane z 3 sekwencji RM, z modelem liniowej regresji wytrenowanym w oparciu o jedną sekwencję RM.	83
5.5	Wyniki oceny procesu gojenia z wykorzystaniem fuzji cech i zbioru treningowego.	86
5.6	Wyniki oceny procesu gojenia z wykorzystaniem fuzji cech i zbioru testowego.	87
5.7	Ocena dryfu krzywej łączącej minima lokalne wartości MAE w macierach błędów z klasą przyjętą jako tydzień rehabilitacji.	88
6.1	Wyniki szkolenia się sieci w paradygmacie end-to-end z wykorzystaniem zbioru treningowego.	99
6.2	Test set results for the tendon healing progression.	100

6.3 Wyniki szkolenia sieci na zbiorze treningowym dla danych z USG. . .	103
6.4 5CV results for the tendon healing progress using end-to-end approach	104
6.5 Porównanie wyników oceny automatycznej bazującej na danych USG i RM na zbiorze testowym.	105
6.6 Wyniki analizy czynników głównych dla 16-tu pomiarów deficytów mięśniowych w warunkach izometrii i izokinetyki. Oznaczono ładunki większe niż 0,7.	106
6.7 Wyniki analizy czynników głównych dla 16-tu pomiarów deficytów mięśniowych, oddzielnie dla pozycji kolana zgiętego i wyprostowanego. Oznaczono ładunki większe niż 0,7.	107
6.8 Komponenty i ich udział, wchodzące w skład nowych zmiennych opisujących biomechanikę.	108
6.9 Rezultaty analizy czynników głównych dla oceny radiologicznej i oceny automatycznej realizowanych z wykorzystaniem 6-ścio parametrycznej ankiety. Oznaczono ładunki większe niż 0,7.	108
6.10 Korelacja między oceną radiologa, a oceną automatyczną. Oznaczone wsp. korelacjami są istotne z $p < 0,05$ dla $N=30$	109
6.11 Korelacja badań biomechanicznych z testem ATRS. Oznaczone wsp. korelacjami są istotne z $p < 0,05$ dla $N=30$	109
6.12 Korelacja badań biomechanicznych i testu ATRS z badaniami radiologicznymi ocenionymi automatycznie i przez radiologa. Oznaczone wsp. korelacjami są istotne z $p < 0,05$ dla $N=30$	109

Rozdział 1

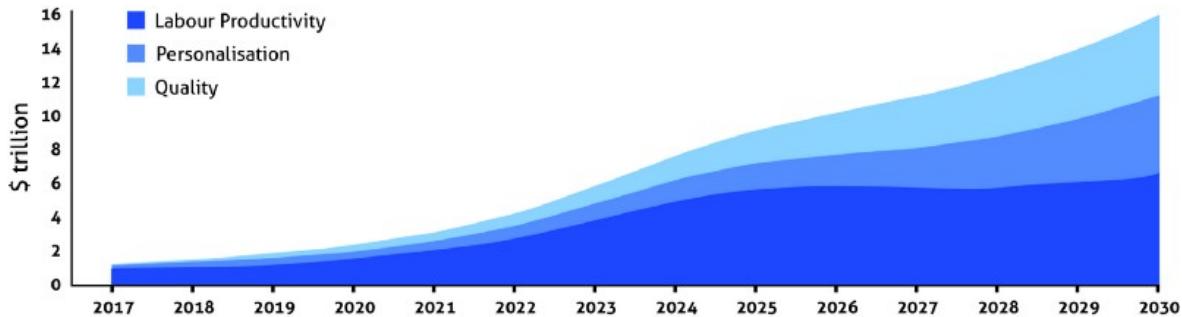
Wstęp

Radiologia zajmuje się obrazowaniem ciała człowieka z wykorzystaniem zjawisk fizycznych umożliwiających nieinwazyjny wgląd wewnętrz organizmu. Z uwagi na szerokie możliwości diagnozowania chorób z wykorzystaniem radiologii, dziedzina ta ma coraz większe znaczenie w medycynie. Z danych przedstawionych przez Siemens Healthineers w 2018 roku wynika, że w ciągu ostatniej dekady liczba skanów z wykorzystaniem Tomografii Komputerowej i Rezonansu Magnetycznego, co roku zwiększa się o 10-12%. Jak wynika z katalogu ambulatoryjnych świadczeń diagnostycznych Narodowego Funduszu Zdrowia, w Polsce, w roku 2018 było to około 1 mln. skanów Rezonansu Magnetycznego i około 4 mln. Tomografii Komputerowych. Do tego należy doliczyć inne modalności jak Rentgen czy Pozytonowa Tomografia Emisyjna. Różne firmy badawcze np. PMR wskazują również, że w Europie, w tym w szczególności w Polsce dokonuje się dużych inwestycji infrastrukturalnych, które dalej implikować będą wzrost liczby obrazowań.

Z drugiej strony, ponownie z danych przedstawionych przez Siemens Healthineers wynika, że wzrost liczby radiologów nie przekracza 3% rocznie. W zestawieniu z dynamiką liczby obrazowań, rezultatem jest zmniejszenie o połowę dostępnego czasu na opis pacjenta i wzrost błędu interpretacji badań nawet o 16,6 punktu procentowego. Z czego 74% wynika ze skutków przemęczenia radiologów i zaburzeń percepcji.

Powysze fakty indukują potrzebę rozwoju narzędzi wspomagających radiologów, w szczególności pozwalających na automatyzację żmudnych, powtarzalnych zadań. Przewidywania wielu firm m.in. prestiżowej agencji PricewaterhouseCoopers (PwC) przewidują, że radiologia zostanie zrewolucjonizowana w najbliższych latach przez rozwiązania do komputerowego wspomagania z emfazą na aplikacje wykorzystujące metody sztucz-

nej inteligencji. Przewidywania te wyrażone w udziale w produkcie krajowym brutto zastosowań metod Sztucznej Inteligencji (SI) można zaobserwować na Rys. 1.1.



Rysunek 1.1: Globalny udział w PKB rozwiązań opartych o SI aplikowanych na rynku Med-Tech.

W pierwszej kolejności motorem zmian mają być zastosowania usprawniające generowanie raportów, następnie personalizacja diagnostyki i narzędzia poprawiające jej jakość np. służące do uzyskania drugiej lub nawet pierwsi opini. W rezultacie czas spędzony przez radiologa na pisanie raportu ma obniżyć się nawet o 90%, nie zmieniając lub polepszając współczynniki błędów diagnostycznych.

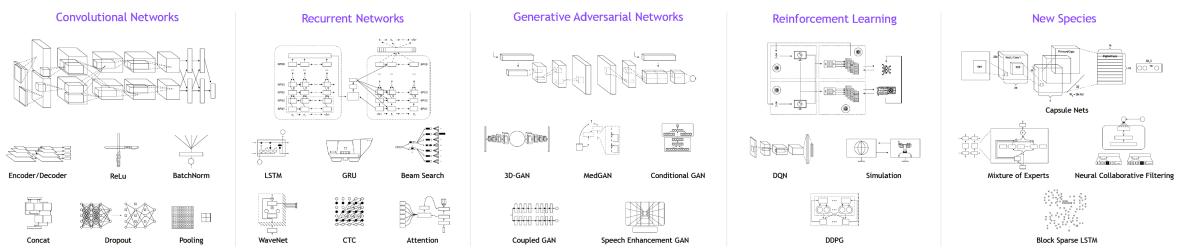
Mówiąc współcześnie o sztucznej inteligencji, a zatem metodach, które według przewidywań mają najbardziej przyczynić się do rewolucji w radiologii, należy rozróżnić dwa jej rodzaje. Pierwsza to tzw. szeroka sztuczna inteligencja (ang. *general Artificial Intelligence*), która docelowo ma odzwierciedlać złożony sposób zachowania się człowieka. Drugi rodzaj nazywany jest wąską Sztuczną Inteligencją (ang. *narrow Artificial Intelligence*), którego metody mają zapewnić możliwie dobry efekt w wykonaniu konkretnego zadania np. klasyfikacji różnego rodzaju tkanek. Przewidywane zmiany w radiologii w najbliższych latach mają być efektem rozwoju metod wąskiej SI. Należy jednak podkreślić, że sam rozwój algorytmów ma już miejsce od ponad pół wieku. Dla przykładu, krótki rys historyczny zawierający lata opracowania wybranych metod można przedstawić następująco: metoda regresji logistycznej – rok 1958, ukryte modele Markova – rok 1960, stochastyczny spadek wzduż gradientu – rok 1960, Maszyna Wektorów Nośnych – rok 1963, algorytm k-najbliższych sąsiadów – rok 1967, wielowarstwowe sieci neuronowe – lata 70-te ubiegłego wieku, algorytm EM – rok 1977, drzewa decyzyjne – rok 1986, Q-learning – rok 1989, lasy losowe – rok 1995 oraz architektury głębokich sieci neuronowych powstające od lat 90-tych.

W szczególności rozwój tych ostatnich został silnie przyspieszony z wykorzystaniem nowych architektur obliczeniowych takich jak akceleratory GPGPU (od ang. *General-*

Rozdział 1: Wstęp

Purpose Computing on Graphics Processing Units), czy TPU (od ang. Tensor Processing Unit). Przede wszystkim uzyskano możliwość szybkiego testowania i optymalizacji architektur zawierających miliony parametrów. W rezultacie dokładność realizacji wybranych praktycznych zadań przez współczesne metody bazujące na głębokich sieciach neuronowych osiągnęła możliwości ludzkiej percepcji.

Wśród głębokich sieci neuronowych wyróżnić można konwolucyjne sieci neuronowe stosowane najczęściej do przetwarzania obrazów, generatywne sieci kontradyktoryjne stosowane najczęściej do generowania danych, rekurencyjne sieci neuronowe stosowane najczęściej do rozpoznawania sekwencji, sieci wykorzystujące uczenie ze wzmacnieniem do modelowania procesów decyzyjnych i szereg nowych klas jak sieci kapsułowe. Mność architektur (zob. Rys. 1.2), powoduje, że na licznych konferencjach czy w publikacjach popularnonaukowych obecny stan rozwoju nazywany jest *eksplozją kambryjską* sieci neuronowych.



Rysunek 1.2: Podział przedstawiający różne rodzaje współczesnych architektur głębokich sieci neuronowych.

Z uwagi na fakt wykorzystania w radiologii w głównej mierze danych obrazowych, to właśnie konwolucyjne sieci neuronowe znajdują tu najczęstsze zastosowanie. Sieci te mają unikalną cechę umożliwiającą przekształcenia informacji obrazowej w numerycznie opisaną klasę przynależności. W szczególności, jak wskazują statystyki przedstawione na konferencji NVIDIA GTC w 2018 roku przez naukowców z niemieckiego centrum badań nad rakiem (DKFZ od niem. *Deutsches Krebsforschungszentrum*), w około 71% przypadków prace dotyczą automatycznego wydzielania struktur anatomicznych, w 12% ich klasyfikacji, w 6% samej detekcji czy poszukiwane zjawisko znajduje się w danych, a w 3% oceny postępu w leczeniu i monitoringu pacjentów. Pozostałe zadania jak np. lokalizacja struktur, czy strukturyzacja danych przyjmują wartości poniżej 2%.

W tej pracy, sieci konwolucyjne zostaną wykorzystane do realizacji metody komputerowego wspomagania radiologa w zadaniu monitorowania pacjentów w trakcie reha-

bilitacji po przebytym urazie całkowitego zerwania ścięgna Achillesa. Zgodnie z przedstawionymi powyżej statystykami, zaprezentowane prace należą do wąskiej grupy ze względu na charakter obecnie realizowane badań, stanowią nowatorskie podejście do przedmiotowego problemu i odpowiadają na istotne problemy w radiologii.

Rozdział 2

Cel i przebieg pracy

Wyniki pracy zostały wykorzystane do realizacji projektu START: "Wykorzystanie autologicznych mezenchymalnych komórek macierzystych w procesie regeneracji rekonstruowanego ścięgna Achillesa", finansowanego przez Narodowe Centrum Badań i Rozwoju z krajowego programu STRATEGMED1. Celem całego projektu było opracowanie nowych nieinwazyjnych technik obrazowania oraz metod monitorowania procesu regeneracji tkanek oraz ocena szybkości gojenia się ścięgna Achillesa. W ramach tych prac powstała koncepcja automatyzacji procesu monitorowania gojenia się ścięgna przy pomocy metod przetwarzania obrazów i sztucznej inteligencji. W szczególności, biorąc pod uwagę budowę głębokich sieci neuronowych i ekstrahowane przez nie cechy można założyć, że będą one skuteczne do oceny procesów patofizjologicznych widocznych w obrazowaniu medycznym, w szczególności regeneracji tkanki miękkiej, co stanowi hipotezę tej pracy.

W ramach weryfikacji hipotezy, za cel główny autor tej pracy postanowił obrać opracowanie automatycznej metody oceny gojenia się ścięgna Achillesa, bazującej na fuzji danych z wykorzystaniem głębokich sieci neuronowych. Jako dane wejściowe postanowiono wykorzystać dane z Rezonansu Magnetycznego (RM), który jest rutynowo stosowany do obrazowania tkanek miękkich i z uwagi na podstawy fizyczne jest też dobrym wyborem do monitorowania gojenia tkanek, zwłaszcza z uwagi na zmiany w ukrwieniu podczas tego procesu. Wedle najlepszej wiedzy autora jest to pierwsza próba automatyzacji oceny gojenia się ścięgna Achillesa widocznego w obrazowaniu medycznym RM. Stąd konieczność zrealizowania celów pobocznych, które stanowią:

1. Wybór efektywnego kosztowo i czasowo protokołu badania bazującego na tech-

nikach obrazowania medycznego, a dokładniej Rezonansu Magnetycznego.

2. Przetestowanie różnego rodzaju podejść związkowych ze szkoleniem głębokich sieci neuronowych.
3. Porównanie wyników oceny nowej metody z wynikami klasyfikacji bazującej na danych z Ultrasonografii.
4. Porównanie wyników oceny nowej metody z oceną funkcjonalną, rutynowo stosowaną do wspomagania rehabilitacji po urazie ścięgna.

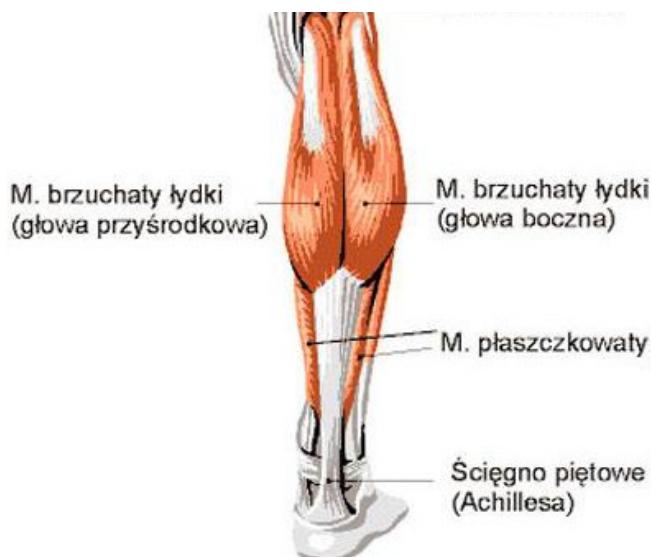
W celu, uporządkowanego i zrozumiałego przedstawienia podłoża oraz wyników badań, w pracy wprowadzono następujący podział na Rozdziały. Po wstępie i opisie celu pracy, jako wprowadzenie, w Rozdziale 3 zostały opisane współczesne metody monitorowania gojenia się ścięgna Achillesa. Rozdział ten rozpoczyna się od omówienia podstaw anatomicznych, biomechanicznych oraz dynamiki procesu gojenia się przedmiotowego ścięgna, a następnie przedstawia szczegółowy opis badań obrazowych i biomechanicznych wykorzystywanych do oceny stanu rehabilitującego się pacjenta. W kolejnym Rozdziale tj. 4, omówione zostały szczegółowo konwolucyjne sieci neuronowe, które posłużyły za rdzeń opracowanego rozwiązania. W szczególności, w opisie uwzględniono zarys historyczny pozwalający zrozumieć etapy rozwoju współczesnych sieci, problemy z budową efektywnych modeli na bazie sieci neuronowych oraz osiągnięcia w tym zakresie wraz z omówieniem sukcesów w zastosowaniach w medycynie. W Rozdziale 5 zaprezentowano nowatorską metodę automatycznej oceny procesu gojenia się ścięgna Achillesa. Co istotne, omówiono unikalny zbiór danych oraz wzorzec odniesienia, dzięki którym możliwa była realizacja przewidzianych badań i walidacji metody. Przedstawiono również eksperymenty wykorzystane do doboru komponentów i parametrów ostatecznego modelu oraz finalny wynik funkcjonowania. W Rozdziale 6, opisano zestawienie wyników metody z wynikami oceny realizowanej przez inne podejścia. W szczególności porównano przedmiotową metodę z metodą uczącą się eksplikytne modelować ocenę radiologa opracowaną również pod kierownictwem autora tej pracy w ramach pobocznych działań. Następnie zaprezentowano zestawienie z wynikami metody działającej na danych z Ultrasonografii, ponownie, opracowanej pod kierownictwem autora tej pracy. Finalnie porównano wyniki z oceną biomechaniczną realizowaną standardowo przez fizjoterapeutów. Całość pracy zakończono podsumowaniem przedstawionym w Rozdziale 7.

Rozdział 3

Monitorowanie procesu gojenia ścięgna Achillesa

3.1 Ścięgno Achillesa

Ścięgno Achillesa, nazywane również ściegnem piętowym, jest największym i zazwyczaj wytrzymującym największe obciążenia ściegkiem występującym w ciele ludzkim. Stanowi wspólne zakończenie mięśnia trójdługiego łydki, w którego skład wchodzą dwie głowy mięśnia brzuchatego i mięsień płaszczkowy. Całość struktury zlokalizowana jest w tylnym, powierzchniowym przedziale łydki co zobrazowano na Rysunku 3.1.



Rysunek 3.1: Lokalizacja mięśnia trójdługiego łydki wraz ze ściegkiem Achillesa.

Procesy patofizjologiczne zachodzące po urazie ściegna są ściśle związane z jego anatomią, biomechaniką, typem urazu i czynnikami jemu sprzyjającymi. Wszystkie te aspekty mają wpływ na możliwość monitorowania procesów gojenia się ściegna i jego wspomagania, dlatego zostaną szczegółowo omówione w kolejnych podsekcjach.

3.1.1 Anatomia

Z obu głów (brzuśców) mięśnia brzuchatego łydki wyrasta jedno szerokie, płaskie pasmo, które jest początkiem części brzuchatej ściegna Achilleса. Następnie ściegno to łączy się z włóknami pochodząymi od mięśnia płaszczkowatego, które układają się stycznie do wcześniej powstałej struktury. Podążając w dół łydki, kształt ściegna Achilleса ulega stopniowemu zwężeniu i zaokrągleniu, aż do punktu o minimalnej szerokości (około 4 cm nad przyczepem dolnym [42]). W rejonie samego przyczepu, znajdującego się na guzie piętowym, ściegno ponownie jest płaskie i szerokie.

Średnia długość ściegna Achilleса to 15 cm (zakres to: 11 – 26 cm). Średnia szerokość w rejonie początku wynosi 6,8 cm (4,5 – 8,6 cm), w rejonie zwężenia to 1,8 cm (1,2 – 2,6 cm), a w miejscu przyczepu dolnego 3,4 cm (2,0 – 4,8 cm) [42, 70].

Ściegno jest zbudowane z tkanki łącznej, a dokładniej mówiąc z tkanki łącznej właściwej zbitej. Składa się w przeważającej części z *istoty międzymiędziorękowej* (nazywanej też *macierzą międzymiędziorękową*) zbudowanej z istoty podstawowej (ang. *ground substance*) oraz włókien. Dopełnienie struktury stanowią komórki takie jak fibroblasty, komórki tuczne, komórki plazmatyczne, histiocity i komórki napływowo.

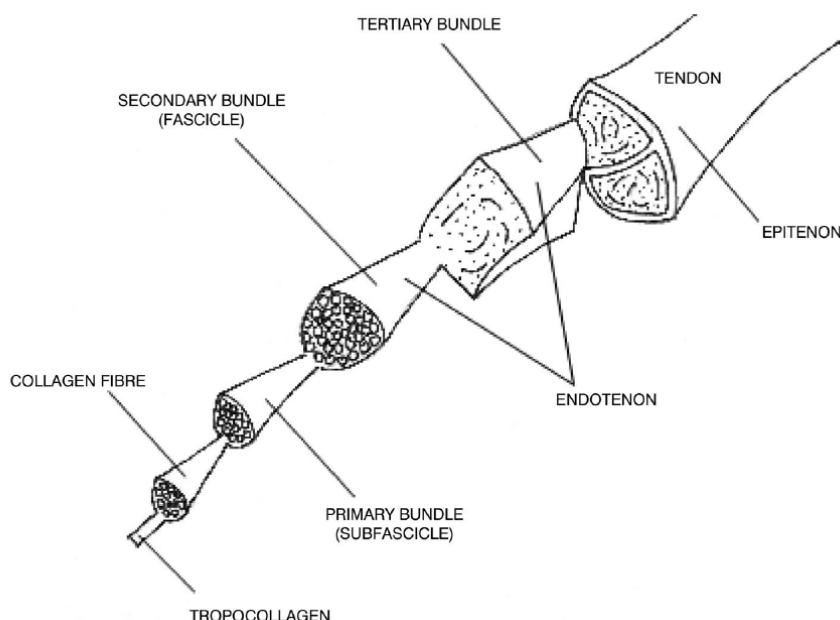
Istota podstawowa jest rodzajem żelu wiążącym duże ilości wody, włókien i komórek. Dokładniej mówiąc pełni ona rolę otoczki zapewniając możliwość transportu wody i substancji odżywczych do wnętrza struktury [116], gdzie znajdują się komórki poukładowane w wąskie pasy leżące pomiędzy włóknami [84]. Wśród włókien można wyróżnić trzy rodzaje: kolagenowe, siateczkowe i sprężyste.

Włókna kolagenowe i siateczkowe są zbudowane z fibrylnego białka – kolagenu, najczęściej występującego białka w organizmie człowieka, stanowiącego około 25% wszystkich białek. Makrocząsteczki kolagenu składają się ze zwiniętych łańcuchów polipeptydowych tworzących helisy. Dotychczas zidentyfikowano 20 typów kolagenu różniących się od siebie szczegółową budową helisy. Kolagen typu I stanowi około 90% masy wszystkich typów włókien i jest podstawowym budulcem włókien kolagenowych tworzących zazwyczaj wiązki o grubości 50–100 μm . Włókna siateczkowe zbudowane

są natomiast w przeważającym stopniu z kolagenu typu III. Pozostałe typy kolagenu występują w znacząco mniejszym stopniu tworząc włókienka kolagenowe i siateczkowe o różnej grubości – od 10 do 300 nm (por. [111]).

Włókna sprężyste występują w postaci sieci i mają średnicę 0,2–10 μm . Zbudowane są z *elastyny*, rozciągliwego i sprężystego białka. Dzięki temu włókna te pod wpływem działania siły zewnętrznej mogą zwiększać swoją długość nawet o 50% (za. [111]).

W strukturze zdrowego ścięgna Achillesa widoczne są ściśle upakowane włókna sprężyste i kolagenowe. Włókna kolagenowe stanowią 70% masy suchego ścięgna i charakteryzuje się hierarchiczną strukturą zobrazowaną na Rys. 3.2.



Rysunek 3.2: Schemat hierarchicznej budowy ścięgna Achillesa.

Wyróżnić można: włókna, wiązki pierwszo, drugo i trzeciorządowe oraz ościegno [116]. W przeciwieństwie do innych ścięgien, ścięgno Achillesa nie posiada pochewki ścięgnistej, lecz jest otoczone ościegiem utworzonym z tkanki łącznej włóknistej. Struktura ta umożliwia ślizg wewnętrznych włókien, jest bogata w naczynia krwionośne i pełniąc funkcje transportowe jest bardzo ważnym elementem w procesie gojenia.

Oprócz otaczającej tkanki łącznej, ścięgno czerpie źródło unaczynienia z brzuśców mięśni brzuchatego i płaszczkowatego oraz z połączenia kostno-ścięgnistego. Najsłabsze unaczynienie występuje na poziomie ok. 4-5 cm powyżej górnego brzegu kości piętowej (por. [27]).

Unerwienie rejonu ściegna Achillesa zapewniają nerw piszczelowy, biegnący wzdłuż całego ściegna, a także nerw łydkowy, który krzyżuje się ze ściegkiem w odległości 8,7-12,4 cm proksymalnie od guza piętowego (por. [27]).

3.1.2 Biomechanika

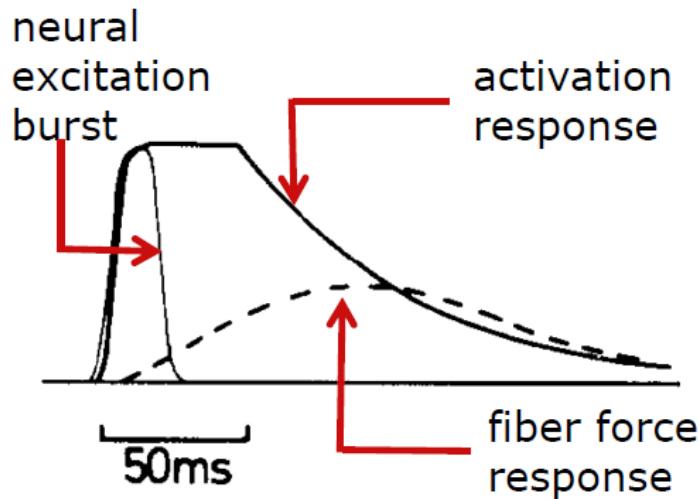
Zadaniem ściegien jest transfer siły mięśniowej do układu szkieletowego. Pod względem mechanicznym ściegno piętowe jest najwytrzymalszym ściegiem całego ustroju. Dla przykładu podczas chodu maksymalne obciążenie ściegna Achillesa wynosi 500 N, przy biegu jest to 9000 N, natomiast podczas wyskoku może sięgać 12000 N, co stanowi równoważność 12–15 krotnej masy ciała. Podobne obciążenia wytrzymuje tylko ściegno właściwe rzepki [122].

Głównym zadaniem ściegna Achillesa w trakcie prawidłowego chodu, biegu, czy wyskoku jest ruch *zginania podeszwowego stopy*, a zatem wyprost powodujący wspięcie na palce. Takie zadanie implikuje zwiększone ryzyko nadmiernego napięcia, a w konsekwencji przeciążenia, rozwoju stanu zapalnego, a nawet uszkodzenia ściegna [122].

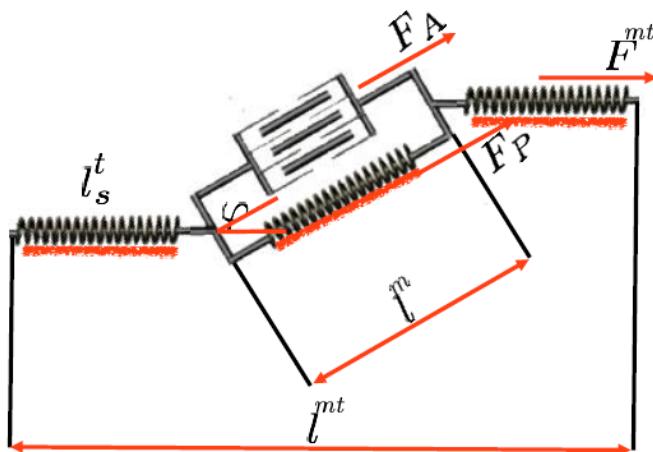
Cały proces ruchu rozpoczyna się od centralnego układu nerwowego skąd wysyłane są impulsy nerwowe. Trafiają one do odpowiednich grup mięśniowych za pośrednictwem *nerwu ruchowego*, struktury przewodzącej impulsy wzmacniające proces skurczu mięśni. Na styku nerwu z mięśniem znajduje się *synapsa nerwowo-mięśniowa* (tzw. *płytką motoneuronalną*). Do niej właśnie na skutek impulsu wydzielany jest neuroprze-kaźnik, *acetylocholina*. W wyniku działania tej substancji następuje dalsze pobudzenie błony komórki mięśniowej i uwolnienie jonów wapnia Ca^{2+} . Cząsteczki te aktywują białka kurczliwe tj. *aktynę* i *miozynę*, a także pośrednio wywołują produkcję w mięśniu *adenozyno-5'-trifosforanu*, *ATP*, czyli substancji zapewniającej dostarczenie odpowiedniej energii chemicznej dla procesu.

Skurcz mięśni szkieletowych trwa 100–300 milisekund. Schemat czasów poszczególnych aktywacji przedstawiono na Rys. 3.3. Jak wcześniej zaznaczono, siła wygenerowana przez skurcz jest przekazywana za pomocą ściegien do układu szkieletowego. Dobrym przybliżeniem tego procesu jest *model Hill'a* [60]. Jego warianty, których przykłady opisane są np. w [100] służą obecnie w wielu rozwiązańach do symulacji pracy układu mięśniowo-szkieleto-wego. Przykład z popularnego rozwiązań OpenSim rozwijanego na Uniwersytecie w Stanford znajduje się na Rys. 3.4.

Na podanym schemacie, F_A oznacza siłę aktywnie wywoływaną skurczem mięśni,



Rysunek 3.3: Schemat generowania skurczu mięśni.



Rysunek 3.4: Schemat modelu typu Hilla stosowany w oprogramowaniu OpenSim.

F_P to pasywna siła odpowiadająca za bezwładność tkanki miękkiej w mięśniu, a F^{mt} to wynikowa siła przekazywana przez ścięgno, która zależy od długości ścięgna l_t i kąta ϕ pomiędzy ścięgkiem a mięśniem.

To właśnie na skutek przekroczenia wartości granicznej siły F^{mt} dochodzi najczęściej do urazu ścięgna Achillesa. Dokładny opis tego problemu wraz z jego implikacjami został przedstawiony w kolejnej podsekcji.

3.1.3 Urazy i czynniki im sprzyjające

Uszkodzenie ściegna Achillea jest jednym z najczęściej występujących urazów układu mięśniowo-szkieletowego. Przykładowo, dla społeczeństwa amerykańskiego, urazy te występują u 18-tu na 100.000 osób rocznie. Dodatkowo ryzyko ponownego zerwania ściegna wynosi 20–40% (zob. [101]).

Można wyróżnić dwa mechanizmy uszkodzenia ściegna Achillea:

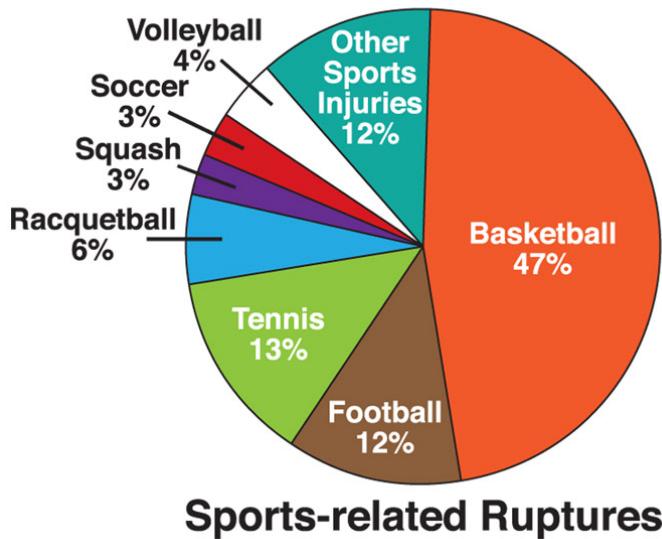
1. uraz bezpośredni – do którego zaliczyć można otwarte urazy spowodowane np. przecięciem ostrym przedmiotem takim jak szkło oraz urazy zamknięte spowodowane nagłym uderzeniem w napięte ściegno.
2. uraz pośredni – znacznie częściej spotykany niż uraz bezpośredni. Powstaje w wyniku nagłego, silnego skurcza mięśnia połączonego często z innymi siłami zewnętrznymi towarzyszącymi np. upadkowi.

O ile uraz bezpośredni jest najczęściej następstwem nieszczęśliwego wypadku, o tyle urazy pośrednie mają swoją przyczynę w rodzaju wykonywanych czynności oraz predispozycji danej osoby.

Najczęściej do tego rodzaju urazów dochodzi podczas uprawiania sportu. Dotyczy to zwłaszcza sportu rekreacyjnego (około 70% przypadków). Aż 75% przypadków przytrafia się mężczyznom w przedziale wiekowym wahającym się od 30 do 50 roku życia. Statystyki te mają związek z obniżeniem poziomu ukrwienia ściegna po 30 roku życia i trybem życia związanym ze sporadycznym, ale nadmiernym obciążaniem ściegna przez osoby niewytrenowane (zob. [122]). Istotne dane zebrane na duńskim społeczeństwie, przeanalizowane w [50], pokazują również, że tendencja przedmiotowych urazów jest wzrostowa i wynika głównie z faktu starzejącego się społeczeństwa oraz zwiększonej liczby zerwań w grupie wiekowej po 50-tym roku życia.

Dyscypliny sportowe, podczas których najczęściej dochodzi do urazu ściegna Achillea to: piłka nożna, siatkówka, tenis, taniec, koszykówka, skoki, biegi, piłka ręczna, aerobic, pływanie, ping-pong, biegi, narciarstwo, balet. Proporcje są naturalnie uzałożnione od popularności danego sportu w konkretnym kręgu ludzi. Dla przykładu na Rys. 3.5 zobrazowano udział poszczególnych dyscyplin w urazach ściegna Achillea dla amerykańskiego społeczeństwa.

W Stanach Zjednoczonych sportem największego ryzyka jest koszykówka, natomiast w Europie jest to piłka nożna (około 60%). Według szacunków przedstawionych w [32]



Rysunek 3.5: Wykres przedstawiający proporcje występowania urazu ścięgna Achillesa w różnych dyscyplinach sportowych na przykładzie amerykańskiego społeczeństwa (na podst. [101]).

uraz ścięgna Achillesa stanowi 5–10% wszystkich urazów atletycznych.

Powszechnie uważa się, że zdrowe ścięgno piętowe jest bardzo ciężko zerwać z uwagi na jego dużą wytrzymałość. Dla przykładu wycinek ścięgna o przekroju 1 cm² jest w stanie utrzymać masę 500–1000 kg [87]. Do urazów dochodzi zatem najczęściej, gdy ścięgno jest zmienione patologicznie. Czynniki powodujące osłabienie ścięgna Achillesa można podzielić na: zewnętrzne i wewnętrzne.

Do czynników wewnętrznych należy zaliczyć:

- mechaniczne uszkodzenie ścięgna – uraz bezpośredni (np. częściowe przecięcie);
- zmiany degeneracyjne – inwolucyjne (wiek powyżej 30 roku życia), na podłożu stanów zapalnych (przewlekłe stany zapalne), związane z mikrourazami;
- nadmierne przeciążenia – nieprawidłowa aktywność fizyczna;
- choroby metaboliczne i układowe – takie jak np. toczeń trzewny, dna moczanowa, nadczynność tarczycy, gruczolaki przysadki mózgowej, reumatoidalne zapalenie stawów, kolagenozy;
- odchylenia anatomiczne i biomechaniczne – anomalie takie jak: piszczel szpotawa, stopa koślawą, stopa płasko-koślawą, niestabilność stawu skokowego, koślawość

lub szpotawość tyłostopia, osłabienie i przykurcz mięśni okołostawowych, ograniczenie ruchomości stawu skokowo-goleniowego i skokowo-piętowego, nierówność kończyn dolnych, nieprawidłowy stereotyp chodu, pogłębane krzywizny kręgosłupa, zespół ciasnoty przedziałów powięziowych w obrębie przedziału tylnego goleni;

- zaburzenia naczyniowe – schorzenia takie jak: hyperlipidemia, cukrzyca lub wywołane nadmiernym paleniem tytoniu;
- dysbalans mięśniowy;
- zaburzenia proprioceptywne;
- niepełne wyleczenie poprzednich obrażeń;
- przyjmowanie kortykosteroidów;

Do czynników zewnętrznych zaliczyć można:

- błędy w metodyce prowadzenia zajęć – np. nadmierny trening, trening jednostronny, trening wszechstronny, z przewagą intensywności nad wytrzymałością czynnościową aparatu więzadłowo-stawowo-mięśniowego, chęć nadrobienia zaległości treningowych;
- nagła zmiana dyscypliny sportowej lub aktywności rekreacyjnej;
- inne takie jak: nieodpowiednia nawierzchnia, niekorzystne warunki pogodowe, nieodpowiednie obuwie, zbyt krótkie przerwy między zawodami lub treningami.

Obecnie, ustalenie jednej optymalnej metody leczenia dla każdego rodzaju uszkodzenia ściegna Achillea nie jest możliwe. Wybór uwarunkowany jest poziomem i stopniem zniszczenia ściegna, stanem zdrowia pacjenta oraz kwalifikacjami, doświadczeniem i możliwościami danego ośrodka. W kolejnej sekcji omówione zostaną współczesne metody leczenia tego schorzenia, natomiast w Rozdziale 5 zaprezentowana zostanie nowa metoda monitorowania procesu gojenia się ściegna Achillea, która może być wykorzystana do optymalizacji rehabilitacji opisanego problemu.

3.1.4 Fazy gojenia, leczenie i rehabilitacja

Proces gojenia się ścięgna Achillesa można podzielić na trzy zachodzące na siebie etapy. W pierwszej kolejności dochodzi do stanu zapalnego. Jest to tzw. *faza zapalna* (ang. *inflammatory phase*), która trwa w pierwszym tygodniu gojenia się. Następnie występuje *faza proliferacji* (ang. *proliferative phase*), trwająca od drugiego do szóstego tygodnia. Cały proces kończy się *fazą przebudowy* (ang. *remodeling phase*), która może trwać nawet do 18 miesiąca po urazie [41, 106, 116, 142].

Faza zapalna zaczyna się od razu po wypadku. W miejscu urazu tworzy się skrzep krwi oraz uwalniane są prozapalne substancje chemiczne, cytokiny, które w znaczącym stopniu odpowiadają za proces aktywacji i migracji komórek zapalnych z otoczenia do miejsca zapalenia [80]. Komórki te, a dokładniej mówiąc leukocyty i makrofagi, rozpoczynają *fagocytozę*, czyli proces mający na celu usunięcie obumarłej tkanki oraz skrzepu [25, 80, 142]. Symultanicznie dochodzi do aktywacji i selekcji *tenocytów* (komórek ścięgna), które rozpoczynają odbudowę włókien ścięgna [142] oraz inicjują drugą fazę całego procesu.

W fazie proliferacji poziom komórek zapalnych zaczyna spadać. Czynniki wzrostu, uwalniane m.in. przez makrofagi, nasilają migrację i rozmnażanie się tenocytów wokół miejsca urazu. Skutkiem tego jest wytworzenie nowych włókien kolagenu typu III. Nowo-powstałe struktury są początkowo losowo zorientowane [25, 41, 142]. Dlatego rozpoczyna się trzecia faza mająca na celu odtworzenie struktury zdrowego ścięgna.

W trakcie trwania ostatniej fazy, nazywanej również *remodelingiem* włókna kolagenu typu III zastępowane są kolagenem typu I. Pod wpływem naprężeń struktura włókien zmienia się i wyrównuje zgodnie z kierunkiem występujących sił [142]. Włókna pracujące niezgodnie z przewidzianym ruchem ulegają zerwaniu i fagocytozie, a te które wspomagają zadany ruch ulegają wzmocnieniu.

Pomimo całego wyżej opisanego procesu, ścięgno wygojone strukturą odbiega od ścięgna zdrowego. Charakteryzuje się nawet dwukrotnie zwiększoną powierzchnią przekroju osiowego oraz podatnością na ponowne zerwania z uwagi na zmniejszoną odporność na obciążenia.

Uraz ścięgna Achillesa można leczyć zachowawczo lub operacyjnie. Zachowawcze leczenie stosuje się najczęściej przy częściowym uszkodzeniu ścięgna lub u osób, które mają przeciwwskazania do leczenia chirurgicznego. Wówczas stosowane jest unieruchamianie kończyny np. poprzez włożenie w but ortopedyczny lub gips. Dodatkowo można

również stosować preparaty z komórek macierzystych i czynników wzrostu [106]. Wybór leczenia zachowawczego wiąże się z długotrwałym, trwającym około 10 tygodni unieruchomieniem kończyny przed przystąpieniem do rehabilitacji.

Leczenie operacyjne stosowane jest najczęściej w przypadku uszkodzenia całkowitego. Tuż po urazie lub po przejściu najintensywniejszych procesów fazy zapalnej wykonuje się szycie ściegna. Stosowane są metody operacji na otwartym ściegnie lub przezskórnie. Zabieg przezskórny polega na zbliżeniu kikutów ściegna specjalnym szwem przeprowadzanym przez małe nacięcia. W teorii metoda ta umożliwia szybsze obciążanie nogi po operacji, skracając okres unieruchomienia i umożliwiając szybszą rehabilitację. W praktyce jednak chirurgom rzadko udaje się uzyskać jakość zszycia porównywalną z metodą otwartą.

Rehabilitacja po urazie również podzielona jest ze względu na fazy gojenia. W pierwszej fazie terapia skupia się wokół ochrony miejsca urazu, odciążenia nogi, redukcji obrzęku i złagodzenia stanu zapalnego. Dokonuje się również zabiegów mających na celu wczesną mobilizację ściegna i przywrócenia ślizgu. Ochrona miejsca urazu może być wykonana przez zastosowanie dwu-częściowej łuski gipsowej lub buta ortopedycznego. Redukcję obrzęku wykonuje się poprzez np. masaż limfatyczny. Stan zapalny łagodzony jest chłodzeniem i okładami oraz lekami, natomiast ślizg przykładowo stymuluje się elektrostymulacją oraz mobilizacją mięśni łydki poprzez zginanie stopy.

W drugiej fazie zabiegi rehabilitacyjne są kontynuacją działań wykonywanych podczas pierwszej fazy. Szczególny nacisk przewidziany jest na przeciwdziałanie powstawania zrostów poprzez pracę nad ślizgiem ściegna oraz mobilizacją tkanek miękkich. W szczególności rehabilitanci poświęcają dużo uwagi na ćwiczenia dotyczące mobilizacji blizny i tkanek okalających. Nasilane są również zabiegi związane z elektrostymulacją celem utrzymania masy mięśniowej oraz wykonywane są ćwiczenia związane ze stabilizacją stawu skokowego związane z ćwiczeniami izometrycznymi mięśni prostujących i pronujących stopę. W tej fazie stosuje się wciąż ochronę miejsca urazu oraz aplikuje się metody redukcji obrzęku.

Trzeci etap tj. przebudowa to najdłuższy z etapów gojenia się tkanek ściegnistych, podczas którego zachodzi najwięcej zmian w strukturze, odporności ściegna oraz jego funkcji. W tym okresie można wyszczególnić następujące cele rehabilitacji:

- przeciwdziałanie zrostom tkanek otaczających Achillesa;
- wzmacnienie mięśni łydki – brzuchatego i płaszczkowatego;

- przywrócenie pełnego zakresu ruchu i kluczowych funkcji ścięgna;
- zwiększenie elastyczności tkanek wokół ścięgna.

Oprócz podstawowych zabiegów ochronnych i fizjoterapii do działań włączane są terapie związane z ćwiczeniami czynnymi z oporem; stymulujące technikę prawidłowego chodu; ćwiczenia wzmacniające obręcz biodrową; ćwiczenia sensomotoryczne w różnych pozycjach ciała oraz ćwiczenia wzmacniające kończyny dolne na niestabilnym podłożu. Powoli wprowadzane są również aktywności sportowe takie jak rower stacjonarny czy stretching. Celem finalnym tego etapu jest powrót do pełnej sprawności.

Dużą nadzieję wiążę się z nowymi technikami bazującymi na zastrzykach z komórek macierzystych. Ciekawą propozycją wydaje się możliwość budowania biodegradowalnych rusztowań (ang. *scaffold*), implementowanych podczas operacji i wspomagających implementację komórek macierzystych i czynników wzrostu w odpowiednie miejsca [30].

Celem oceny tego rodzaju nowatorskich podejść jak również z uwagi na konieczność optymalizacji istniejących metod rehabilitacji i leczenia urazów ścięgna Achillesa potrzebne jest monitorowanie procesu gojenia. Najczęściej stosowane techniki zostaną omówione w kolejnych sekcjach.

3.2 Zastosowanie rezonansu magnetycznego

Obserwację gojącej się struktury ścięgna Achillesa umożliwiają techniki stosujące różnego rodzaju fale przenikające ciało ludzkie. Sposób w jaki propaguje się fala tworząc sygnał zależy od własności związanych z ośrodkiem propagacji oraz parametrów fali. Do najistotniejszych należą wyrażana w hercach (w skr. Hz) częstotliwość, czyli liczba pełnych cykli drgań w ciągu sekundy, amplituda tj. maksymalna wartość sygnału falowego oraz faza określająca w której części cyklu znajduje się sygnał. W technikach obrazowania medycznego wykorzystywana jest interpretacja zdarzeń wynikających ze zmian tych parametrów oraz reagującej materii.

Pierwszą opisaną w tej pracy metodą obrazowania stosowaną do monitorowania gojenia się ścięgna Achillesa jest obrazowanie z wykorzystaniem *Jądrowego Rezonansu Magnetycznego* nazywanego dla uproszczenia *Rezonansem Magnetycznym*, w skr. RM. Metoda ta bazuje na odpowiednim odczycie reakcji jąder atomów na stałe pole magnetyczne generowane przez magnesy (zwykle nadprzewodzące), zmienne pole magnetyczne ze składową prostopadłą do osi stałego pola (generowane w dedykowanej

do tego zadania cewce) oraz zmienne pole lokalne wytwarzane przez sąsiednie jądra atomowe.

Zrozumienie własności jądra atomowego i reakcji na magnetyzację było możliwe dzięki badaniom naukowym z początku XX wieku. Zapewniły one praktyczną możliwość wykorzystania procesów bazujących na zachowaniach obiektów o bardzo małych masach i wymiarach takich jak cząstki elementarne. Jako przykłady można wskazać prace Alberta Einsteina dotyczące ruchów Browna oraz efektu fotoelektrycznego; opis funkcji falowej zaproponowany przez Erwina Shrodingera; badania Nielsa Bohra, które doprowadziły do utworzenia utylitarnego modelu atomu zawierającego poziomy energetyczne; wyprowadzenie reguły nieoznaczoności przez Werner Heisenberga, czy opisanie przez Wolfganga Pauliego konfiguracji cząstek elementarnych w atomie.

W dużej mierze na bazie powyższych prac zdefiniowana została teoria mechaniki kwantowej umożliwiająca między innymi pogłębienie wiedzy na temat magnetycznej natury składników jądra atomowego tj. *protonów* i *neutronów*. Pierwszy ważny przełom nastąpił w 1933 r., kiedy to Frisch i Stern w [49] opisali moment magnetycznego protonu. Rok później, w [29] Breit i Rabi zaproponowali metodę rezonansową obserwacji własności magnetycznych jąder atomowych. W 1946 praca ta została udoskonalona i wykorzystana przez Felix Blocha do badania cząstek w wodzie i parafinie [26]. W końcu dzięki pracom m.in. Lautenbur'a [75] i Mansfield'a [86] rozwinęły się metody lokalizacji przestrzennej powyższych zmian własności magnetycznych, co umożliwiło utylitarną prezentację wizualną wyników.

Opisując zagadnienie dokładniej, należy zwrócić uwagę na główną własność jądra atomowego umożliwiającą działanie RM, czyli *spin*. Tylko jądra o nieparzystej liczbie masowej, posiadają niezerowy spin w stanie podstawowym. Liczba masowa jest sumą liczb protonów i neutronów (por. [94]). Dla przykładu jądro wodoru składające się z 1 protonu spełnia powyższy warunek.

Umieszczone w polu magnetycznym jądra atomowe posiadające spin wykonują ruch wirowy zwany *precesją*. Częstotliwość tej precesji jest proporcjonalna do natężenia pola magnetycznego i opisana jest *równaniem Larmora* 3.1:

$$\omega_0 = \gamma * B_0, \quad (3.1)$$

gdzie ω_0 to częstotliwość kołowa protonów, $f_l = \omega_0/2\pi$ to tzw. częstotliwość Larmora, γ to współczynnik żyromagnetyczny właściwy danemu jądru atomowemu, a B_0 to wartość

indukcji danego pola magnetycznego. Dla przykładu częstotliwość Larmora dla jądra wodoru, gdzie $\gamma = 42,57 \text{ MHz/T}$ dla pola $1,5\text{T}$ równa jest $63,8 \text{ MHz}$.

Jądra atomowe wirujące z częstotliwością Larmora na skutek silnego pola magnetycznego ulegają swego rodzaju uporządkowaniu, tworząc rotujące w przybliżony sposób wektory magnetyzacji. Dodatkowo istotne jest, że jądra o spinach ustawionych równolegle mają mniejsze energie, a o ustawionych antyrównolegle większe. Jest możliwe zatem, na skutek sekwencji impulsów fal radiowych, aby spiny jąder uległy odchyleniu. W przeciętnym wycinku ciała ludzkiego jest znacząco więcej jąder o ustawieniu równoległym niż antyrównoległym. Przykładowo, w stosunkowo niewielkim w porównaniu z wymiarami typowo obrazowanych struktur anatomicznych kawałku objętości $1 \times 1 \times 5 \text{ mm}$, jest to liczba około 2×10^{15} cząstek w polu o indukcji $1,5\text{T}$.

Sygnal RM pochodzi od różnicy magnetyzacji jąder ustawionych równolegle i antyrównolegle (różnicę tę oznaczamy jako M_0 – tj. całkowita magnetyzacja). Wyjściowo M_0 ma większą wartość zgodną ze zwrotem pola magnetycznego, czyli tzw. *magnetyzację podłużną*. Kierunek prostopadły do magnetyzacji podłużnej nazywany jest *magnetyzacją poprzeczną*.

Zmianę M_0 wykonuje się przy pomocy odpowiednio sterowanej cewki nadającej impulsy o częstotliwości fal radiowych (tzw. *impulsy RF* – od ang. *Radio Frequency*). Impuls RF jest krótkotrwały i ma taką samą częstotliwość, co częstotliwość precesji. Dzięki temu zachodzi zjawisko *rezonansu magnetycznego*, czyli wymiany energii między jądrami atomowymi, a falą radiową. Cząstka pochłania wówczas energię *fotonu*, tj. nośnika oddziaływania elektromagnetycznego, i uzyskuje wyższy poziom energetyczny, odchylając wektor magnetyzacji i zaburzając tym samym uporządkowanie. Następuje zmniejszenie M_0 w kierunku podłużnym oraz wzrost magnetyzacji poprzecznej. Po wyłączeniu impulsu RF magnetyzacja wraca do wartości początkowej.

Czas potrzebny na zwiększenie magnetyzacji podłużnej do stanu z przed działania impulsu nazywany jest czasem *relaksacji podłużnej* lub *relaksacji spin – sieć*. Z kolei czas potrzebny na obniżenie magnetyzacji poprzecznej nazywany jest czasem *relaksacji poprzecznej* lub *relaksacji spin – spin*. Relaksacja spin–spin ma przede wszystkim znaczenie w kontekście oddziaływań z sąsiednimi jądrami i występującą na skutek tego dyfuzją fazy obniżającą wartość magnetyzacji poprzecznej. Istotnymi parametrami obu powyższych procesów są odpowiednio stałe $T1$ i $T2$, określające dynamikę omówionych zmian. Z reguły $T2 \leq T1$.

Lokalizacja zmian magnetyzacji w poszczególnych *wokselach* (tj. wycinkach prze-

strzeni 3D o określonych wymiarach) odbywa się przy pomocy celowanego różnicowania fazy i częstotliwości wektorów magnetyzacji. Operacja ta wykonywana jest przy pomocy sygnału z dodatkowych cewek gradientowych. Wartości w warstwie, w której dokonywany jest pomiar kodowane są przy pomocy nałożenia na pole B_0 pola gradientowego G_{zz} , co umożliwia wzbudzenie jąder atomowych w tych wokselach, dla których współrzędna z wyraża się wzorem:

$$\Delta z = \frac{\Delta\omega}{\gamma G_{zz}}, \quad (3.2)$$

gdzie $\Delta\omega$ to szerokość widmowa pobudzającego jądra impulsu RF. Współrzędna y kodowana jest przy pomocy zróżnicowania fazy, rotujących w różnych obszarach wektorów magnetyzacji, na skutek działania pola gradientowego G_{yz} — tzw. *gradientu kodowania fazy*. Zróżnicowanie fazy wyraża się wzorem:

$$\phi = \gamma G_{yz} y T_y, \quad (3.3)$$

gdzie T_y oznacza czas włączenia pola gradientowego G_{yz} . Ostatecznie, współrzędna x kodowana jest w wyniku zróżnicowania częstości precesji wektorów magnetyzacji poprzez włączanie pola gradientowego G_{xz} (zob. [5]).

Nieprzetworzona informacja z wokseli o współrzędnych x, y, z nazywana jest *przestrzenią K* i koduje ona składniki częstotliwości sygnału RM. Do transformacji z przestrzeni częstotliwości na przestrzeń obrazu używana jest operacja matematyczna zwana *odwrotną transformacją Fouriera*. Przykład takiego działania można znaleźć w [14].

Główną informacją jaką otrzymujemy za pomocą obrazowania metodą jądrowego rezonansu magnetycznego jest rozkład gęstości jąder atomowych o niezerowym spinie. Jądra wodoru mają największy spośród jąder atomowych współczynnik żyromagnetyczny oraz występują w znaczającej liczbie, w prawie każdej z tkanek ludzkich (por. [94]). Dlatego ich rozkład jest najczęściej mierzony. Jedynie w szczególnych przypadkach stosuje się obrazowanie z wykorzystaniem częstotliwości rezonansu odpowiednich dla innych pierwiastków takich jak fosfor [124], sód [123] czy węgiel [44].

Dodatkowo, obserwacja zmiany sygnału RM w czasie, często dostarcza istotnych danych na temat właściwości fizyko-chemicznych struktur anatomicznych. W szczególności uzyskanie informacji na temat parametrów $T1$ i $T2$ zapewnia możliwość różnicowania tkanek. Z uwagi na ten fakt, dla uwydatnienia interesujących informacji, stosuje się szereg ustaleń urządzeń do rezonansu magnetycznego pogrupowanych w

sekwencje RM oraz ich podgrupy nazywane w tej pracy *modalnościami*. Do badań opisanych w dalszej części tego manuskryptu zostało użytych 7 sekwencji, w tym jedna z 4 modalnościami. Zostaną one pokrótko scharakteryzowane poniżej:

- *T1 zależne* – zależność sygnału rezonansu magnetycznego MR_s w tej sekwencji opisana jest w następujący sposób:

$$MR_s \sim \gamma_{pd} * [1 - e^{-TR/T1}], \quad (3.4)$$

gdzie γ_{pd} to gęstość protonowa tkanki, a $T1$ odpowiada momentowi, gdy wartość magnetyzacji podłużnej wynosi 63% wartości początkowej. Czas ten nie jest znany, stąd użytkownik określa *czas repetycji* TR rzędu spodziewanego czasu $T1$.

- *T2 zależne* – zależność sygnału rezonansu magnetycznego MR_s w tej sekwencji opisana jest w następujący sposób:

$$MR_s \sim \gamma_{pd} * e^{-TE/T2}. \quad (3.5)$$

$T2$ odpowiada momentowi, gdy wartość magnetyzacji poprzecznej wynosi 37% wartości początkowej. Podobnie jak w poprzedniej sekwencji, czas ten nie jest znany, więc użytkownik określa parametr TE , *czas echa*, rzędu czasu $T2$. Sekwencja T2 zależne charakteryzuje się odmienną skalą szarości niż T1 zależne.

- *PD* (od ang. *Proton Density*) – obraz rekonstruowany na podstawie sygnału RM przy bardzo krótkim TE i bardzo długim TR . Wówczas sygnał jest wprost proporcjonalny do γ_{pd} :

$$MR_s \sim \gamma_{pd}. \quad (3.6)$$

- *T2 mapping* – w tej sekwencji wartość sygnału RM zależna od $T2$ (zob. 3.5) mierzona jest dla 8 TE . Pozwala to na wyliczenie czasu $T2$ (zob. [103]).
- *T2 * GRE* (od ang. *Gradient Echo*) – jest to przykład *sekwencji gradientowych*, gdzie spin obracany jest za pomocą impulsów RF o mniejszej niż 90° . Zabieg ten używany jest w celu przyspieszenia pomiaru kosztem jakości sygnału. Stąd też oznaczenie czasu $T2^* < T2$.
- *T2 * GRE TE_MIN* (od ang. *Minimal Time Echo*) – obraz rekonstruowany jest na podstawie sygnału RM dla $T2^*$ przy minimalnym czasie TE i tylko dla fragmentu przestrzeni K. Zabieg ten skracza znaczco czas badania.

- 3D FSPGR (od ang. *Three-dimensional Fast Spoiled Gradient Echo*) – jest to szybka sekwencja gradientowa z usuwaną (ang. *spoiled*) pozostałością magnetyzacji poprzecznej. W wyniku sekwencji impulsów RF otrzymywane są 4 możliwości:
 - In Phase Ideal – sygnał mierzony w czasie, gdy protony należące do tłuszcza i wody są w zgodnej fazie;
 - Out Phase Ideal – sygnał mierzony w czasie, gdy protony należące do tłuszcza i wody są w antyfazie;
 - Fat Ideal – sygnał mierzony przy maksymalnej wartości sygnału od protonów tłuszcza i minimalnej od protonów wody;
 - Water Ideal – sygnał mierzony przy maksymalnej wartości sygnału od protonów wody i minimalnej od protonów tłuszcza.

Powyższe możliwości są możliwe do realizacji, gdyż w sekwencji 3D FSPGR sygnał RM odczytywany jest przy dwóch różnych czasach *TE*. Przy czym czasy *TE* są tak dobrane, aby sygnały od wody i tłuszcza wzmacniały się (In Phase Ideal) lub osłabiały (Out Phase Ideal). Jest to możliwe gdyż przy zadanym polu magnetycznym częstotliwość Larmora wodoru w wodzie jest minimalnie inna niż w tłuszczach (fakt ten nazywany jest również przesunięciem chemicznym i jest wykorzystywany w spektroskopii NMR [79]). Modalności Water Ideal i Fat Ideal są odpowiednią kombinacją liniową dwóch wcześniejszych – taką aby zminimalizować wpływ tłuszcza (Water Ideal) lub wody (Fat Ideal) na sygnał RM.

Rezonans magnetyczny może być wykorzystany do monitorowania gojenia się ścięgien i więzadeł, gdyż tkanki miękkie człowieka zawierają dużo atomów wodoru. W kontekście ścięgna Achillesa RM w szczególności pozwala radiologom na obserwację: ciągłości ścięgna w płaszczyźnie strzałkowej (ang. *sagittal plane*); uszkodzeń śródścięgnistych objawiających się przerwaniem w naturalnym warkoczu ułożonym z tkanek; pogrubienia ścięgna ocenianego w przekrojach osiowych (ang. *axial planes*); jednorodności, wrzecionowatości i innych nieregularności kształtu; ostrości ścięgna i jego rozgraniczenia od tkanek otaczających wraz ze zmianami występującymi na brzegach; obrzęku w okalających ścięgno tkankach; jednorodności charakteryzującej się podobieństwem przekrojów sąsiednich oraz liczby ewentualnych zrostów.

Ograniczenia związane ze stosowaniem RM wynikają głównie z oddziaływaniami silnego pola magnetycznego na przedmioty z nim reagujące np. ferromagnetyki przykładowo

zawarte w odłamkach metalowych; urządzenia elektryczne typu wszechsepiony stymulator serca lub duże kawałki metalu jak endoprotezy, które mogą się rozgrzać na skutek indukowanych prądów. Wrażliwość pacjenta na hałas lub skłonności klaustrofobiczne również są wskazaniami do wykluczenia z badania. Ponadto urządzenie do rezonansu magnetycznego jest bardzo drogie, nieprzenośne i wymagające specjalnego otoczenia pracy. W celu wyeliminowania tych ograniczeń można posłużyć się inną metodą obrazowania medycznego opisaną w kolejnej sekcji.

3.3 Zastosowanie ultrasonografii

Kolejną z metod obrazowania medycznego jest *Ultrasonografia*, w skr. *USG* (ang. *Ultrasonography, US*). Bazuje ona na efektach związanych z propagacją w tkankach *ultradźwięków*, tj. fal akustycznych o częstotliwościach powyżej 20 kHz. W praktyce, z uwagi na parametry związane z rozdzielczością obrazu, są to fale o częstotliwości rzędu kilku MHz, w szczególnych przypadkach kilkudziesięciu.

Propagacja fal akustycznych w przyrodzie była tematem rozważań myślicieli takich jak Pitagoras, Arystoteles czy Galileusz, którzy ugruntowali pole badań pod kolejne osiągnięcia matematyczno-inżynieryjne. W tej kwestii, do jednego z przełomów doszło w 1822 roku, kiedy to szwajcarski inżynier Daniel Colladen oraz matematyk Charles-Francois Sturm wyznaczyli przybliżoną prędkość rozchodzenia się fali akustycznej w wodzie. Badanie wykonano na Jeziorze Genewskim symultanicznie mierząc czas jaki potrzebny był dźwiękowi podwodnego wystrzału i sygnałowi dzwonka rozchodzącego się w powietrzu do przebycia drogi pomiędzy dwoma łódkami oddalonymi o 10 mil. Wyliczona wartość wyniosła wówczas 1435 m/s nie różniąc się znacząco od dzisiaj przyjmowanej estymacji równej 1480 m/s. Analogicznie, stosując nowoczesne techniki, wyliczane są obecnie prędkości rozchodzenia się fali w innych ośrodkach takich jak powietrze, tkanki miękkie, kości itp.

58 lat po eksperymencie na Jeziorze Genewskim, w 1880 roku bracia Curie opisali w [64] *efekt piezoelektryczny*, czyli zjawisko polegające na pojawienniu się ładunku elektrycznego pod wpływem naprężeń mechanicznych w krysztale o anizotropowej budowie, takiej jak np. kwarc. W przypadku odwrotnym, przyłożenie napięcia do odpowiedniego kryształu generuje drgania (tzw. *odwrotny efekt piezoelektryczny*).

Efekty te są wykorzystywane w głowicy aparatu usg, przyrządzie do generowania i odbierania ultradźwięków. Przykładowo, polaryzowanie kryształu piezoelektrycznego

3.3 Zastosowanie ultraso

krótkim impulsem elektrycznym $\sigma_1(t)$ pobudza go do drgań na własnej częstotliwości rezonansowej. Zakładając, że ów kryształ ma kształt walca o grubości $d=0,64$ mm, to będzie stanowił rezonator półfalowy, w którym wystąpi drganie rezonansowe o długości fali $\lambda = 2d$ czyli $\lambda = 1,28$ mm. Jeżeli wykonany jest z tytanianu baru, dla którego prędkość propagacji drgań wynosi $c=4460$ m/s, to częstotliwość drgań własnych tego kryształu wyniesie:

$$f = \frac{c}{\lambda} = \frac{4460 \text{m/s}}{1,28 \text{mm}} = 3,5 \text{MHz}. \quad (3.7)$$

Odwrotnie, powracająca fala tzw. *echo* wygeneruje impuls elektryczny $\sigma_2(t)$ w skutek drgań wywołanych w kryształku. Echo jest naturalnie falą różniącą się od sygnału nadawanego, a zmiany te są w przeważającym stopniu efektem zjawisk takich jak: *odbicia, załamania, dyfrakcja, rozpraszanie i pochłanianie* (zob. [85]).

Zjawiska te zależą od częstotliwości fali, która propaguje się w ośrodku o pewnej *impedancji akustycznej ośrodka* Z wyrażanej jako:

$$Z = \rho c = \sqrt{\epsilon \rho}, \quad (3.8)$$

gdzie c , to prędkość rozchodzenia się fali, ρ to gęstość ośrodka, a ϵ to *moduł odkształcalności objętościowej*, tj. parametr opisujący jak zmieni się objętość ośrodka pod danym ciśnieniem. Parametry wybranych ośrodków zestawiono w Tabeli 3.6.

Ośrodek	Moduł Sprężystości [kg/m/s ² × 10 ⁻⁸]	Gęstość [kg/m ³]	Prędkość [m/s]	Impedancia akustyczna- [kg/m ² /s × 10 ⁻⁸]
Powietrze	0,0000134	1,2	330	0,0004
Woda (20°C)	2,19	1 000	1 480	1,48
Rtęć	28,5	13 600	1 450	20,0
Tkanki miękkie	2,51	1 060	1 540	1,63
Tkanka tłuszczowa	2,0	952	1 450	1,38
Wątroba	2,54	1 060	1 550	1,64
Mięśnie wzdużnie	2,74	1 080	1 592	1,70
Mięśnie poprzecz.	2,80	1 080	1 610	1,74
Mózg	2,40	994	1 550	1,55
Śledziona	2,59	1 045	1 578	1,64
Krew	2,47	1 057	1 057	1,62
Kość	32,0	1 912	4 080	7,8
Piuka	0,169	400	650	0,26
Soczewka oka	3,03	1 142	1 620	1,85
Ciało szkliste	2,31	1 000	1 520	1,52
Ciało wodnistre	2,25	1 000	1 500	1,50
Kwarc	88,0	2 650	5 750	15,2
Tytanian baru	107,0	5 400	4 460	24,0
Olej mineralny	2,1	969	1 477	1,43

Rysunek 3.6: Parametry ośrodków często mierzonych w badaniach USG.

Dla przykładu, do zjawiska załamania lub odbicia dochodzi kiedy fala pada na granice dwóch ośrodków o różnych impedancjach akustycznych Z_1 i Z_2 . Dla fali prostopadłej zależność ta opisana jest następująco:

$$R = \frac{I_r}{I_0} = \left(\frac{Z_1 - Z_2}{Z_1 + Z_2} \right)^2, \quad (3.9)$$

gdzie I_r to natężenie fali padającej, a I_0 odbitej. Natomiast R , czyli współczynnik odbicia, jest parametrem, który rośnie wraz ze wzrostem kąta odchylenia od kierunku prostopadłego, aż do całkowitego odbicia.

Z kolei do rozpraszania bądź pochłaniania fali dochodzi kiedy to fala pokonuje daną drogę w ośrodku o pewnej Z , co zapisywane jest następująco:

$$I = I_0 e^{-\gamma x}, \quad (3.10)$$

gdzie γ to współczynnik osłabienia zależny od Z , a x to droga przebyta przez falę. Efekt ten można korygować poprzez dobór odpowiedniego I_0 lub mierząc czas powrotu i na tej podstawie wyliczając spodziewane osłabienie. Zabieg ten pozwala obrazować jednocześnie struktury położone w różnych odległościach.

Analiza amplitudy i częstotliwości sygnału nadanego i echa umożliwia rekonstrukcję obrazu USG. W przypadku najczęściej stosowanych w praktyce rekonstrukcji przekrojów dwuwymiarowych (tzw. *tryb B*) współczesny tor budowania prezentacji wizualnej (tzw. *beamforming*) wygląda następująco:

1. Główica ultradźwiękowa emisuje impulsy w postaci wąskiej wiązki w ścisłe określonym kierunku. Wiązka jest efektem interferencji sygnału z N przetworników zawierających kryształy piezoelektryczne.
2. Echa z danego kierunku pozwalały na obliczenie pojedynczego promienia akustycznego, który jest ilorazem charakterystyk nadawanego i odbieranego sygnału.
3. Wszystkie promienie, których we współczesnych aparatach może być do kilkuset (zob. [7]), służą do formowania obrazu, który tworzony jest we *współrzędnych biegunowych* (r, θ) w przypadku główic mechanicznych sektorowych, wieloelementowych convex czy fazowych lub we współrzędnych prostokątnych (x, y) w przypadku główic mechanicznych lub wieloelementowych liniowych¹.

¹szczegółowy opis główic USG i ich charakterystyk można znaleźć w [15]

Tryb B umożliwia również wizualizację obrazów dynamicznych. Przykładowo jeżeli na obraz składa się 400 promieni i każdy odsłuchiwanym jest do głębokości 15 cm, to czas gromadzenia danych dla ośrodka o średnim $c=1500 \text{ m/s}$ wynosi $2 \frac{2 \times 15}{1500 \text{ m/s}} \times 400 = 0,08 \text{ s}$, czyli 12 obrazów na sekundę. Częstotliwość tę można zwiększać, zmniejszając liczbę promieni lub głębokość obserwacji.

Innym często stosowanym trybem rekonstrukcji obrazu jest *tryb D* bazujący na *efekcie Dopplera*, do którego dochodzi w przypadku przechodzenia fali przez ośrodek przesuwający się względem głowicy. Zmienia się wówczas częstotliwość fali, co wyrażone jest następującym wzorem:

$$f_r = 2f_o \frac{v}{c} \cos(\theta), \quad (3.11)$$

gdzie f_r to zmiana częstotliwości fali nadawanej f_o , zależna od kąta θ pomiędzy falą i ośrodkiem poruszającym się z prędkościami rozchodzenia się fali w obu ośrodkach tj. v i c . Dla przykładu, im większa prędkość komórek przesuwających się w monitorowanym ciele pacjenta, tym większa jest f_r . Dlatego tryb D z powodzeniem jest wykorzystywany np. do monitorowania przepływu krwi w dużych naczyniach takich jak tętnice.

Rozwinięciem trybu D jest tryb *Power D* (od ang. *Power Doppler*). Gdzie zamiast przesunięcia częstotliwości interpretowana jest moc sygnału Dopplerowskiego. Za pomocą odpowiedniego parametru nazywanego *gain* można wzmacniać sygnał Power D uwidaczniając np. informacje pochodzące od przepływów krwi z niewielką prędkością w charakteryzujących się małą średnicą naczyniach krwionośnych. Jest to również możliwe, gdyż tryb Power D ma nawet trzykrotnie większą czułość niż standardowy tryb D [21]. Wynika to z faktu, że Power D nie uwzględnia informacji o prędkości czy kierunku przepływów, a jest sumą wszystkich możliwych przesunięć częstotliwości w zadanym fragmencie obrazowanego obiektu.

W kontekście ścięgna Achillesa tryb Power D może służyć do oceny unaczynienia ścięgna w kolejnych etapach gojenia, które jak wiadomo z sekcji 3.1.4 zmienia się w czasie. Tryb B natomiast może być użyteczny do obrazowania struktury tkanek miękkich. W praktyce wykorzystywana jest zwyczajna możliwość zobrazowania ukierunkowania struktur włókien ścięgnistych na podstawie czego radiolog może wnioskować o fazie gojenia. Składowa czasowa jest interesująca z perspektywy m.in. fizjoterapeuty oceniającego ślizg w ścięgnie przy wykonywaniu odpowiednich ruchów np. zginania podeszwowo-grzbietowego stopy.

W porównaniu do rezonansu magnetycznego, w kwestii ograniczeń, należy zwrócić uwagę na fakt, że fale akustyczne używane w USG nie propagują się dobrze przez kości

i gazy. Zastosowanie do oceny struktur umiejscowionych w otoczeniu lub składających się w większości z takich ośrodków np. odpowiednio mózg lub płuca znajduje wówczas RM.

Z użyciem RM możliwe jest uzyskanie obrazów o lepszej jakości detali. Sam czas badania jest natomiast dłuższy i w większości przypadków niemożliwe jest obrazowanie w czasie rzeczywistym, co z kolei jest naturalne dla techniki USG.

W wymiarze finansowym istotny jest fakt, że aparat do USG kosztuje średnio 10 razy mniej niż aparatura do RM. Jak wspomniano jednak, z uwagi na jakość detali, uzyskiwane obrazy są trudniejsze do interpretacji, stąd trudności w wyszkoleniu kadry.

W omawianych aspektach na znaczące zmiany mogą wpływać nowe rozwiązania w warstwie sprzętowej i oprogramowania. Do pierwszej grupy należy zaliczyć zastąpienie przetworników z piezoelektrykami, przetwornikami budowanymi w technologii MEMS np. cMUT, czy pMUT (zob. [10]) oraz układy pozwalające przetwarzać surowy sygnał ultradźwiękowy (zob. [16]). Dzięki przetwornikom MEMS można wytworzyć cały układ generujący drgania w krzemie w jednym procesie technologicznym razem z dedykowanym układem do zadanej aplikacji (ang. *Application-Specific Integrated Circuit* w skr. *ASIC*). Takie podejście znaczowo redukuje koszty oraz implikuje możliwość miniaturyzacji urządzeń.

Do drugiej grupy należą algorytmy sztucznej inteligencji pozwalające wydobyć i zinterpretować interesującą informację z niskiej jakości obrazów. Przykładowo w [36] algorytmy sztucznej inteligencji zostały użyte do określenia orientacji włókien mięśniowych, a w [13] do segmentacji komór serca w czasie rzeczywistym.

Obiecujący jest też rozwój metod charakterystyki tkanki na podstawie ultrasonografii (ang. Ultrasonography Tissue Characterization, w skr. *UTC*). W 2003 r., w [134], Bakker et. al. przedstawili pracę pokazującą w jakim stopniu obraz USG jest mieszaną echa związanego z budową tkanki, a w jakim z *interferencji*, czyli nakładania się fal. W zależności od stabilności echa można zatem wnioskować czy obserwowana struktura składa się głównie z dużych, stałych struktur wywołujących stabilne echo, czy np. płynów lub małych włókienek powodujących zmienne interferencje. Jako referencji dla studiowanego echa używa się badań histologicznych (zob. [133]). Na tej podstawie sklasyfikowano 4 różne rodzaje echa i jak zasygnalizowano w pracach [135] i [59] informacja o proporcjach występowania tych rodzajów może być użyta do oceny struktury ścięgna Achillesa. Metoda ta nie została jednak jeszcze w pełni zwalidowana i możliwości wnioskowania na jej podstawie są wciąż niejasne (zob. [59]).

USG i inne techniki obrazowania medyczne nie są jedynymi metodami oceny gojenia się ścięgna Achillesa. W kolejnej sekcji zostały opisane techniki oceny funkcji ścięgna, które samodzielnie jak i w połączeniu z analizą obrazową stanowią wartościową информацию diagnostyczną.

3.4 Zastosowanie badań biomechanicznych

W poprzednich sekcjach zostały opisane dwie najczęściej wykorzystywane obrazowe metody monitorowania procesu gojenia się ścięgna Achillesa. Komplementarnie, podczas rehabilitacji może być również wykonana *ocena funkcjonalna*, technika weryfikująca w jakim stopniu dany element (tkanka, narząd, organizm) może realizować swoje zadania. W przypadku monitorowania gojenia się ścięgna Achillesa najczęściej w tym celu stosuje się *ocenę biomechaniki*, czyli badania pozwalające wnioskować na temat właściwości mechanicznych elementów składowych organizmów żywych.

Najbardziej zaawansowane i dokładne metody oceny biomechaniki stosowane współcześnie możliwe są do wykonania przy użyciu urządzeń pomiarowych takich jak:

- *Komputerowa analiza ruchu* (ang. *Motion Capture*) – narzędzie wykorzystujące systemy czujników do zapisu informacji o zmianach położenia obiektu rejestrowanego np. pacjenta. Do wiodących rozwiązań należą zaliczyć systemy firmy Vicon [17], czy BTS [3].
- *Płyty dynamometryczne* (ang. *Force Plates*) – narzędzie wykorzystywane do pomiaru sił reakcji podłoża w trzech prostopadłych płaszczyznach. Dzięki temu można określić sumaryczny udział mięśni w generowaniu sił odpowiadających za balans ciała, ruch w danym kierunku oraz przecistawianie się sile grawitacji. Do wiodących rozwiązań należą płyty firmy Kistler [8].
- *Elektromiografia*, w skr. EMG (ang. *Electromyography*) – narzędzie do pomiaru pobudzeń poszczególnych grup mięśniowych podczas ruchu. Wykorzystywane jest m.in. do określenia rozkładu sił zmierzonych przez płyty dynamometryczne na poszczególne mięśnie.

Synchronizacja danych z powyższych urządzeń umożliwia konstrukcję modeli układu mięśniowo-szkieletowego i symulacje funkcji poszczególnych grup mięśniowych oraz ścięgien przy zadanych problemach.

Danymi stosowanymi do uszczegółowienia takich modeli są np.: wymiary poszczególnych segmentów ciała (goleń, udo, tors) tzw. pomiary antropometryczne; maksymalne siły izometryczne mierzone z użyciem systemów takich jak Biodes; geometria kości mierzonych np. z pomocą tomografii komputerowej (w skr. TK) ew. RM; lokalizacja przyczepów mięśniowych określanych przy pomocy RM lub USG; środki masy poszczególnych segmentów określanych np. przy użyciu badania *DXA* (od ang. *Double X Ray Absorption*); skład włókien mięśniowych widocznych w USG.

Z uwagi na dużą liczbę możliwych do zmierzenia parametrów, ich integracja odbywa się w modelach komputerowych zaimplementowanych w różnego rodzaju oprogramowaniu do symulacji biomechanicznych. Do najczęściej używanych modeli należą opisane w [65] Gait2392 i Gait 2354 oraz obecnie najbardziej złożony – *AnyBody Full Body Model* [22].

Historia komputerowo wspomaganego, kompleksowego modelowania biomechaniki ruchu sięga wczesnych lat 90-tych ubiegłego wieku, kiedy to Delp i Loan przedstawili oprogramowanie SIMM [39]. Obecnie SIMM jak również inne oprogramowania komercyjne takie jak: Visual 3D (Cmotion Inc.) [18], Anybody (Anybody Technology) [2], czy Adams (MSC Software Corp.) [1], dostarczają narzędzi do wartościowych symulacji np. chodu [121], biegu [55] jak również konsekwencji różnych zabiegów chirurgicznych [52] i chorób [115]. Istnieją również narzędzia otwartha, do których należą szeroko wykorzystywany OpenSim [38] rozwijany na Uniwersytecie w Stanford, czy też Human Motion [128] wywodzący się z instytutu badawczego RIKEN w Japonii.

Powyżej przedstawione kompleksowe badania w praktyce realizowane są rzadko z uwagi na wysokie koszty. Dla przykładu w Polsce, ośrodki wyposażone w sprzęt pomiarowy takiej klasy to np. Instytut "Pomnik – Centrum Zdrowia Dziecka", Warszawski Uniwersytet Medyczny, Akademia Wychowania Fizycznego imienia Józefa Piłsudskiego, czy komercyjna placówka Fizjofit w Gliwicach. Żeby obniżyć koszty badania stosuje się wybiórcze podejście i selekcję parametrów pomiarowych uznanych przez ekspertów dziedzinowych za wystarczające do analizy zadanego problemu. Dla przykładu badania stosowane do oceny biomechaniki ścięgna Achillesa w placówce Carolina Medical Center (gdzie realizowane były badania wykorzystywane w tej pracy) składają się z następujących pomiarów (zob. [106]):

1. Pomiar *ATRS* (od ang. *Achilles Tendon Total Rupture Score*) – oceniany jest w skali od 0 do 100 [96] poziom ograniczenia, z którymi pacjenci borykają się w następstwie urazu.

3.4 Zastosowanie badań Biomedycznych Monitorowanie procesu gojenia ścięgna Achillesa

2. Pomiar stabilograficzny na platformie dynamometrycznej – mierzone są wychylenia środka ciężkości. Pacjent ma za zadanie utrzymanie równowagi na niestabilnym podłożu. Badanie realizowane jest boso z oczami otwartymi. Wykonane są dwie próby po 30 sekund kolejno na prawej i lewej kończynie dolnej na dynamometrycznej platformie dynamometrycznej Bidex Balance System. Wyniki zostają porównywane między kończynami.
3. Pomiar stabilograficzny na platformie statycznej – mierzona jest droga wychylenia środka masy pacjenta w trakcie stania jednonóż na platformie dynamometrycznej, statycznej.
4. Pomiar sił reakcji na ścieżce podometrycznej – mierzony jest rozkład sił nacisków podeszwowej strony stóp na podłożu (jedynie w kierunku prostopadłym do podłożu). Pomiar wykonywany jest podczas stania swobodnego, wspięć na palce oraz przysiadu boso bez odrywania pięt. Dokonywana jest również analiza chodu (3 przejścia) i biegu (5 przebiegnięć). Boso oraz w obuwiu sportowym. Na podstawie sił reakcji wyliczane są parametry: rotacja podudzia [deg], długość kroków [cm], udział fazy podparcia [%], udział fazy przenoszenia [%], maksymalna siła na pięcie [N] oraz maksymalna siła na palcach [N].
5. Pomiar skoczności i mocy (tzw. siły dynamicznej) kończyn dolnych – mierzona jest moc maksymalna P_{max} i średnia P_m , maksymalna wysokość uniesienia h_{max} i obniżenia k środka masy ciała przed odbiciem. Wykonywane są wyskoki pionowe z miejsca na platformie dynamometrycznej. Realizowane są dwie próby obunóż oraz na prawej i lewej kończynie dolnej w obuwiu sportowym. W celu pełnego zaangażowania kończyn dolnych pacjent podczas badania trzyma ręce na biodrach.
6. Pomiary momentów sił mięśni stawu skokowego – mierzone są maksymalne wartości momentu siły mięśni zginaczy podeszwowych i grzbietowych stawu skokowego [Nm] oraz deficyt pomiędzy operowaną i zdrową kończyną dolną [%]. Momenty sił mierzone są w dwóch pozycjach tj. z wyprostowanym oraz zgiętym do 50 stopni stawem kolanowym [98]. Pomiar realizowany jest w warunkach izometrii i izokinetyki w trzech prędkościach kątowych $60^\circ/s$ (5 powtórzeń), $120^\circ/s$ (8 powtórzeń) oraz $180^\circ/s$ (10 powtórzeń) przy wykorzystaniu urządzenia Humac Norm (USA). Przed badaniem osoba odbywa 5 minutową rozgrzewkę na steperze.

Wymienione wyżej badania zostały określone przez ortopedów i fizjoterapeutów jako wystarczające do oceny przywracania funkcji ściegna Achillesa po rekonstrukcji.

Wraz z oceną strukturalną realizowaną poprzez badania obrazowe i wiedzą ekspercką informacja tak zgromadzona może służyć do subiektywnego monitorowania procesu gojenia się ściegna. Do skutecznej obiektywizacji tego procesu potrzebne są jednak dodatkowe metody bazujące na agregacji ilościowych współczynników i automatycznym wnioskowaniu na ich podstawie. Do tej grupy należą algorytmy sztucznej inteligencji opisane w kolejnym rozdziale.

Rozdział 4

Konwolucyjne sieci neuronowe

Konwolucyjne sieci neuronowe (ang. *Convolutional Neural Networks*, CNN) są biologicznie inspirowanymi sztucznyimi sieciami neuronowymi. Należą do zbioru *głębokich sieci neuronowych* (ang. *Deep Neural Networks*, DNN), które z kolei są podzbiorem *systemów uczących się* (ang. *Machine Learning Systems*) tj. algorytmów nie wymagających explicite programowania do ustalenia swoich parametrów.

Pierwsze matematyczne formalizmy dotyczące CNN zostały zaproponowane już w latach 40-tych XX wieku, natomiast fundamentalne inspiracje biologiczne dały badania nad korą wzrokową kotów Hubel'a i Wiesel'a z lat 50-tych i 60-tych (zob. [140]). M.in. dzięki pracom tych neurofizjologów ustalono, że kora wzrokowa zawiera złożone układy komórek, które odpowiadają za przetwarzanie informacji z wybranych regionów pola widzenia, sumarycznie pokrywając je w całości. Komórki kory wzrokowej działają zatem jak lokalne filtry przestrzeni wejściowej zaprojektowane, tak aby wydobyć istotne informacje z naturalnych obrazów. Dla przykładu reagują na orientację linii, kształty i kolory.

W odróżnieniu od naturalnych struktur biologicznych, sieci konwolucyjne operują przeważnie na reprezentacji cyfrowej obrazu. Zapisywana jest ona w postaci tablicy liczb tzw. *macierzy* o współrzędnych (x,y) , gdzie x to kolumna macierzy, a y wiersz. W przypadku obrazów trójwymiarowych dochodzi jeszcze składowa z , a zamiast macierzy użyty jest wówczas *tensor*, czyli uogólniona reprezentacja.

W obrazach cyfrowych dla każdego punktu (x,y) , tzw. *piksela*, kodowana jest informacja o wartości funkcji obrazowej $I(x,y)$. Na tej podstawie dzielimy obrazy na:

- binarne – gdzie kodowane są jedynie dwie możliwe wartości I (0 lub 1);

- monochromatyczne – gdzie kodowana jest informacja o natężeniu jednej barwy (najczęściej są to odcienie szarości lub brązu tzw. *sepia*);
- kolorowe – gdzie kodowane są wartości natężenia składowych koloru.

Do zakodowania informacji o wartości funkcji I w obrazie binarnym potrzebny jest jeden bit na punkt. Odcienie obrazu monochromatycznego lub pojedyncze składowe punktu obrazu kolorowego kodowane są przy pomocy 8–16 bitów, co uzależnione jest od zakresu wartości występujących w danych lub precyzji wymaganej do obliczeń. Przykładowo, obrazy medyczne przetwarzane w tej pracy należą do grupy obrazów monochromatycznych i są zapisywane w reprezentacji 16 bitowej z uwagi na zakres wartości występujących w sygnale RM.

W sieciach konwolucyjnych procesy zachodzące w korze wzrokowej są modelowane przy użyciu odpowiednio dobranych parametrów i funkcji służących do ekstrakcji istotnych informacji obrazowych i ich przetwarzania. Parametry wykorzystywane do wstępnej ekstrakcji informacji grupowane są w *filtre*, które realizują funkcję *splotu* maski K z kolejnymi fragmentami funkcji I . K jest najczęściej macierzą kwadratową o wymiarze N , gdzie zazwyczaj $N = 1, 3, 5, 7, 9, 11$ z uwagi na rosnącą wraz z N złożonością obliczeniową. Równanie splotu dyskretnego można zapisać następująco:

$$I'(x, y) = \sum_{j=-n}^n \sum_{i=-k}^k I(x - j, y - i) K(j, i), \quad (4.1)$$

gdzie I' jest nową funkcją obrazową powstałą po filtracji, a i i j to kolejne współrzędne maski filtru w odniesieniu do jego punktu centralnego. Dla wartości brzegowych podczas realizacji splotu, współrzędne poza sygnałem przyjmują z reguły wartość 0. Rzadziej stosuje się inne metody takie jak: odbicie wartości I obrazu poza jego granicami; powtórzenie wartości I obrazu bez odbicia; modyfikację wymiaru maski filtru na brzegu obrazu, tak by nie wychodziła poza jego granice.

W kontekście sieci konwolucyjnych, pojedynczy splot filtru z obrazem wejściowym nazywany jest *cechą* (zob. [109]). W ostatnich latach powstało wiele algorytmów do wizualizacji cech umożliwiających wydajną analizę wyników działania sieci konwolucyjnych np.: Saliency Maps [117], GradCam [113] czy CNNVIS [83]. Zasada działania tych algorytmów bazuje na selekcji cech, które mają zasadniczy wpływ na wynik końcowy. Grupy cech obliczane na tym samym obszarze obrazu wejściowego nazywane są *mapami cech* (ang. *feature maps*). Zazwyczaj mapy cech zawierają hierarchicznie upo-

rządkiowaną strukturę składającą się z cech od najniższego do najwyższego poziomu komplikacji (np. od charakterystycznych skupisk pikseli, przez obiekty składowe do finalnego kształtu). Do cech najniższego rzędu zaliczyć można:

- *krawędzie* – ciągi punktów o gwałtownych zmianach $I(x,y)$;
- *rogi* – przecięcia dwóch krawędzi;
- *grzbiety* lub *doliny* – odpowiednio lokalne maksimum lub minimum funkcji I ;
- *skupiska* (ang. *blobs*) – obszary jednorodnych wartości funkcji I , różniących się znacząco od najbliższego otoczenia;
- *tekstury* – charakterystyczne, przestrzenne ułożenie wartości funkcji I w powtarzające się wzory.

Wszystkie cechy wyższego rzędu są kombinacją powyższych składowych podstawowych tworząc bardziej skomplikowane struktury odpowiadające charakterystycznym obiektom znajdującym się w obrazach wejściowych.

Na podstawie cech, na końcu toru przetwarzania danych z wykorzystaniem sieci neuronowych, realizowane jest wnioskowanie, gdzie w zależności od problemu można wyszczególnić następujące zadania:

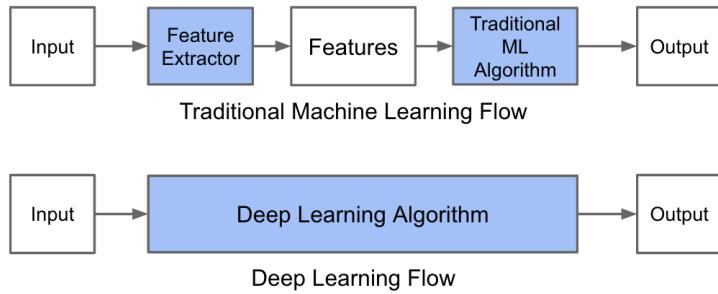
- *segmentację* – podział obrazu na spójne fragmenty, najczęściej wiążące się z wyodrębnieniem *obiektu* z *tła* na podstawie ustalonego progu lub miary np.:

$$\begin{aligned} I(x, y) \geq T_p &\Rightarrow (x, y) \in "obiekt" \\ I(x, y) < T_p &\Rightarrow (x, y) \in "tlo", \end{aligned}$$

przy czym metody ustalania wartości progowej T_p lub wielu wartości T_p^1, \dots, T_p^n są obszarem szerokich badań (zob. [19]);

- *klasyfikację* – przyporządkowanie obiektu do odpowiedniej klasy (np. tkanka zdrowa lub patologiczna);
- *detekcję* – binarne rozróżnienie traktujące o tym czy obiekt znajduje się w obrazie czy nie;
- *śledzenie* – detekcja lub też klasyfikacja obiektów w kolejnych krokach czasowych.

Jeszcze do niedawna tor przetwarzania danych w większości badań wykorzystujących uczenie się maszyn, operacje wyliczenia cech i wnioskowanie końcowego zawierał w dwóch osobnych krokach¹. Porównanie tego schematu z obecnie funkcjonującym podejściem realizowanym w algorytmach głębokiego uczenia się przedstawiono na Rys. 4.1. W nowym podejściu zarówno ekstrakcja cech jak i ostateczne wnioskowanie na ich pod-



Rysunek 4.1: Porównanie schematów przetwarzania danych z wykorzystaniem metod głębokiego uczenia się i innych algorytmów.

stawie realizowane jest w jednym kroku, co nazywane jest paradygmatem *end-to-end learning*. Na przestrzeni lat wprowadzano kolejne składowe, które finalnie utworzyły obecnie funkcjonujące podejście. W kontekście poznania tych komponentów, w kolejnej sekcji omówiono dokładniej zarys historyczny przedstawiający ewolucję nowego paradygmatu.

4.1 Zarys historyczny

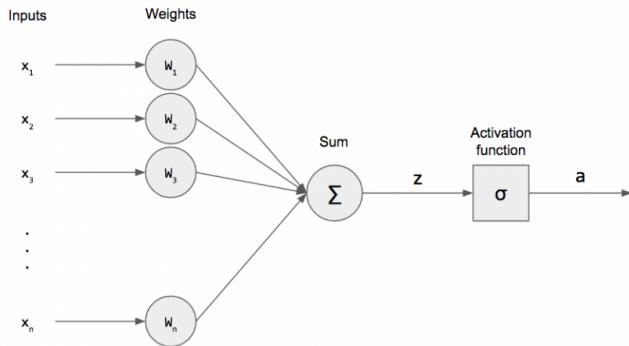
Pierwszy formalny model neuronu został zaproponowany przez Warrena McCulloch i Waltera Pitts w roku 1943 (zob. [92]). Była to bramka logiczna, której wyjście stawało się aktywne w momencie, gdy liczba aktywnych wejść przekroczyła pewien zdefiniowany próg. Taka zależność sygnału wyjściowego y od sygnałów wejściowych $x_1 \dots x_n$ została potem nazwana *funkcją aktywacji neuronu*, którą zapisujemy jako:

$$y = f(x_1, x_2, \dots, x_n) \quad (4.2)$$

W modelu neuronu McCulloch-Pitts można było modyfikować parametr progu, nie istniała natomiast możliwość uczenia się takiej architektury. Problem ten rozwiązano

¹Są wyjątki od tej reguły jak np. [114], inkrementalny algorytm SVM.

w 1957 proponując sztuczną sieć neuronową zawierającą wiele neuronów z ważonymi połączeniami między sobą (zob. [105]). Sieć nazwano perceptronem, co było implikacją zamiłowania jego twórcy Franka Rosenblatta do aplikacji związanych z percepcją, zwłaszcza mowy czy pisma. Schemat sieci pokazano na Rys. 4.2.



Rysunek 4.2: Topologia perceptronu.

Zastosowanie dodatkowej jednowymiarowej tablicy współczynników, czyli *wektora wag* $w = (w_1, \dots, w_n)$ dało możliwość uczenia się poprzez adaptacyjną zmianę wartości poszczególnych jego elementów. W modelu Rosenblatta zastosowano ponadto progową funkcję aktywacji z progiem T_a :

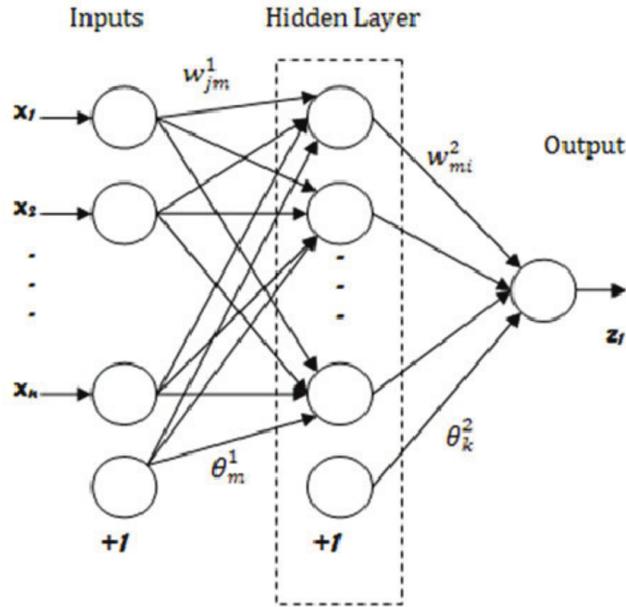
$$y(z) = \begin{cases} 0, & n < T_a, \\ 1, & n \geq T_a, \end{cases} \quad (4.3)$$

gdzie z to suma ważona wyjść poszczególnych neuronów:

$$z = \sum_{i=1}^n w_i x_i \quad (4.4)$$

Dla przykładu, w dwuwymiarowej przestrzeni sygnałów wejściowych x_1, x_2 działanie perceptronu można opisać jako wyliczenie funkcji liniowej rozdzielającej obserwacje o_1 od o_2 . W trzech wymiarach będzie to płaszczyzna, a w n -wymiarach, hiperplaszczyna. Z uwagi na swój liniowy charakter perceptron proponowany przez Rosenblatta miał wiele ograniczeń, które trafnie sformułowali w 1969 roku Marvin Minsky i Seymour Papert w książce *Perceptrons* (zob. [91]). Autorzy opublikowali listę problemów, których nie można było rozwiązać z użyciem perceptronu m.in. do najszerzej dyskutowanych należał przykład związany z brakiem możliwości modelowania funkcji *XOR*, aktywującej wyjście przy aktywnym jednym z dwóch wejść.

Po latach intensywnych prac, część z opisanych przez Minsky-Papert problemów udało się rozwiązać za sprawą pracy Davida Rumelharta, Geoffa Hintona i Ronalda Williams, którzy opublikowali w 1986 roku pracę [107], traktującą o perceptronach wielowarstwowych. Schemat takiej sieci zaprezentowano na Rys. 4.3



Rysunek 4.3: Topologia perceptronu wielowarstwowego.

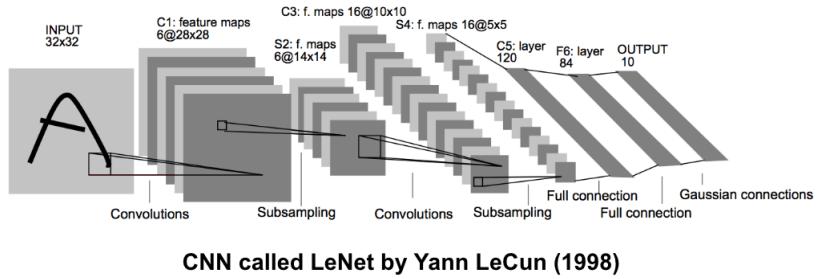
Spośród najważniejszych innowacji wprowadzonych w perceptronie wielowarstwowym wyszczególnić można zastosowanie w praktyce nowego algorytmu uczenia się sieci, który został opisany w kolejnej sekcji jak również nowej, sigmoidalnej funkcji aktywacji:

$$y(x) = \frac{1}{1 + e^{-x}} \quad (4.5)$$

Z wykorzystaniem sieci wielowarstwowych możliwe stało się modelowanie funkcji XOR jak i innych problemów nieliniowych o praktycznym wymiarze.

Kolejny przełom nastąpił w 1989 roku kiedy to Yann LeCunn, uczeń Geoffa Hintona, zaprezentował swoje wyniki dotyczące klasyfikacji odręcznego pisma z użyciem sieci wielowarstwowych (zob. [78]). Finalnie, badania te doprowadziły do przedstawienia w 1998 roku pierwszej sieci konwolucyjnej nazwanej LeNet (zob. [77]). Architekturę tej sieci przedstawiono na Rys. 4.4.

Sieć składała się z 7 warstw i zawierała około 60,000 parametrów. Oryginalnie, sygnał wejściowy sieci stanowił obrazek o wymiarach 32×32 . W architekturze LeNet



Rysunek 4.4: Topologia sieci LeNet.

zaobserwować można było dwie podstawowe składowe współczesnych sieci konwolucyjnych wykorzystywanych do klasyfikacji obrazów tj.:

- *ekstraktor cech* – część zawierająca m.in. filtry, służące do automatycznej ekstrakcji wektora cech w .
- *klasyfikator* – część wykorzystywana do zadania wnioskowania końcowego na podstawie w .

Przy użyciu takiej architektury możliwe stało się zastosowanie paradygmatu end-to-end learning rozumianego w tym kontekście jako znalezienie możliwie dobrej transformacji, która surowe obrazy przekształca bezpośrednio w ostateczną klasyfikację. Dokładny opis szkolenia się sieci konwolucyjnych i problemów z tym związanych został przedstawiony w kolejnej sekcji.

4.2 Szkolenie głębokich sieci neuronowych

Większość algorytmów szkolenia głębokich sieci neuronowych obejmuje zadanie *optymalizacji*, rozumiane jako minimalizację, bądź maksymalizację *funkcji celu* $f(x)$ przez zmianę x . W literaturze można też znaleźć inne nazwy funkcji celu takie jak *kryterium*, *funkcja kosztów*, *funkcja strat* lub *funkcja błędów* (por. [53]).

Podczas zadania optymalizacji bardzo często wykorzystuje się *pochodną funkcji* oznaczaną jako $f'(x)$ lub $\frac{\delta y}{\delta x}$, gdyż niesie ona informacje o nachyleniu funkcji w punkcie x . W praktyce funkcje celu są wielowymiarowe dlatego wykorzystywane są *pochodne cząstkowe* informujące o nachyleniu w poszczególnych wymiarach. Wektor zawierający pochodne cząstkowe funkcji nazywany jest *gradientem* i oznaczany jest jako $\nabla f(x)$.

W praktyce wykorzystywane są również: macierz pochodnych cząstkowych tzw. *macierz Jacobiego* oznaczana jako J oraz macierz *drugich pochodnych* cząstkowych (tj. pochodnych pochodnych) tzw. *macierz Hessego* oznaczana jako H .

Z uwagi na szybkie znajdowanie lokalnych minimów funkcji celu, to właśnie metody optymalizacji bazujące na wartości gradientu są najczęściej używane w szkoleniu głębokich sieci neuronowych. Inne metody, niegradientowe (zob. [130]), przeważnie nie są w tym kontekście wystarczająco efektywne.

Znalezienie lokalnego minimum zwykle nie jest równoważne z otrzymaniem najlepszego możliwego rozwiązania, jednak w praktyce uznawane jest za satysfakcjonujące. Dla potwierdzenia można przytoczyć następujące fakty wynikające z wielu badań (por. [34]):

1. Dla sieci neuronowych o dużych rozmiarach większość lokalnych minimów charakteryzuje się podobnymi wartościami, przekładającymi się na porównywalny efekt wnioskowania końcowego.
2. Prawdopodobieństwo znalezienia lokalnego minimum, którego implikacją będą niezadowalające rezultaty wnioskowania przy użyciu sieci maleje wraz ze wzrostem rozmiaru sieci.
3. Próba znalezienia globalnego minimum bardzo często prowadzi do problemu nadmiernego dopasowania, omówionego dokładniej w sekcji 4.2.1.

Podsumowując, metody gradientowe są wydajne obliczeniowo i prowadzą do znalezienia wielu satysfakcjonujących rozwiązań, które mogą być wykorzystane do rozwiązania praktycznych problemów.

Zasada działania metod gradientowych bazuje na obieraniu w kolejnych krokach iteracji następujących wartości funkcji f , przesuwając się w kierunku spadku gradientu:

$$x' = x - \epsilon \nabla f(x), \quad (4.6)$$

gdzie ϵ to *szkość uczenia się*, parametr określający wielkość kroku.

Funkcja celu w przypadku praktycznych zadań optymalizacyjnych, wykorzystujących głębokie uczenie się jest *funkcją złożoną*, a zatem efekt jej działania jest równoważny operacjom wykonywanym przez kilka lub więcej funkcji po kolei. W takich

sytuacjach do obliczeń spadku gradientu wykorzystywane są funkcje, których pochodne są znane.

Aplikuje się je do tzw. *reguły łańcuchowej*. Przypuśćmy, że $v = g(p)$ i $u = f(g(v)) = f(q)$, gdzie p i q to wektory. Wówczas regułę łańcuchową można zapisać jako:

$$\frac{\delta u}{\delta p_i} = \sum_j \frac{\delta u}{\delta q_j} \frac{\delta q_j}{\delta p_i}, \quad (4.7)$$

co w zapisie wektorowym równoważne jest z równaniem:

$$\nabla_x z = \left(\frac{\delta q}{\delta p} \right)^T \nabla_q u, \quad (4.8)$$

gdzie $\frac{\delta q}{\delta p}$ to macierz Jacobiego.

Regułę łańcuchową zapisaną w 4.8 prosto uogólnia się do zmiennych tensorowych (zob. [53], str. 205) i stosuje się w różnych meta-algorytmach służących do szkolenia sieci. Przykładem jest *algorytm propagacji wstecznej*, który oblicza regułę łańcuchową w wydajnej kolejności stosując działania w grafie takim jak topologia perceptronu wielowarstwowego (zob. [53]).

Proces szkolenia się sieci ma na celu najlepsze możliwe przybliżenie docelowej klasyfikacji bazując na danych przykładach, czyli *zbiorze uczącym* U . Algorytmy optymalizacyjne używane do szkolenia głębokich sieci neuronowych zazwyczaj działają pośrednio, optymalizując pewną miarę wydajności P , która jest zdefiniowana na *zbiorze testowym*, zawierającym przykłady inne niż w U . Często bierze się również pod uwagę jeszcze dodatkowy podzbiór, rozłączny z U i T – *zbiór walidacyjny*, który ma pomóc w wyborze najlepszych algorytmów i wartości parametrów wpływających na jakość szkolenia sieci.

W procesie szkolenia zmniejszana jest wartość funkcji kosztów f w oparciu o U , a celem jest poprawa P . Algorytmy optymalizacyjne wykorzystujące cały zbiór U do liczenia wartości gradientu nazywane są *pakietowymi* lub *deterministycznymi*, gdyż przetwarzają jednocześnie wszystkie przykłady szkoleniowe. Te, które używają jednego przykładu na raz są nazywane *stochastycznymi*. Natomiast w praktyce przy szkoleniu głębokich sieci stosowane są algorytmy *minipakietowe*, wykorzystujące więcej niż jeden przykład, ale mniej niż cały zbiór. Z reguły są to liczby z przedziału 8–256. Takie podejście zapewnia kompromis między szybkością obliczeń i dokładnością estymacji wartości gradientu.

Przykładem algorytmu minipakietowego jest *stochastyczny spadek gradientu* (ang.

stochastic gradient descent, SGD). Bazuje on na założeniu, że estymację gradientu można otrzymać wyliczając średnią gradientu z minipakietu m przykładów. Kolejne kroki algorytmu można zapisać następująco:

1. Wybierz wartość parametru szybkości uczenia się ϵ_k .
2. Wybierz minipakiet złożony z m przykładów ze zbioru szkoleniowego.
3. Oblicz estymację gradientu $g = \frac{1}{m} \nabla \sum_{i=1}^m L(x_i, y_i, f)$, gdzie L to funkcja strat dla jednego przykładu o wejściowej wartości próbki x_i i oczekiwany wyjściu y_i .
4. Zastosuj aktualizację wartości funkcji celu równą $\epsilon_k g$.
5. Jeżeli kryterium stopu nie zostało spełnione wróć do kroku 2.

Kryterium stopu jest najczęściej określone liczbą iteracji lub satysfakcjonującą wartością funkcji f . Kwestia optymalnego wyboru parametru ϵ_k zależy od problemu i najczęściej stosowane są metody empiryczne, przy czym zazwyczaj ϵ_k maleje wraz ze zbliżaniem się do satysfakcjonującego rozwiązania.

Parametry algorytmów szkoleniowych nazywane są *hiperparametrami*, gdyż nie są wyznaczane bezpośrednio w procedurze uczenia się. Strategia nadawania hiperparametrom wartości początkowych jest silnie dyskutowana w literaturze (zob. [69]) i jej kompleksowy opis wykracza poza zakres tej pracy. Warto jednak wspomnieć o algorytmach z adaptacyjną szybkością uczenia się, gdyż jest to jeden z najtrudniejszych do ustawienia hiperparametrów, a jednocześnie bardzo istotny. Są to m.in.:

- Adaptive Gradient Algorithm (AdaGrad) [45] – wykorzystywany do indywidualnej adaptacji szybkości uczenia się wszystkich parametrów modelu, skalując ją odwrotnie proporcjonalnie do pierwiastka kwadratowego sumy wszystkich historycznych kwadratów gradientów.
- Root Mean Square Propagation (RMSProp) [112] – modyfikacja AdaGrad, w której zamiast akumulacji gradientu wykorzystuje się wykładniczo ważoną ruchomą średnią z gradientu.
- Adaptive moments algorithm (Adam) [68] – W porównaniu z RMSProp, Adam poza momentem pierwszego rzędu (tj. średnią) wykorzystuje również moment drugiego rzędu (tj. *wariancję*). Dokładniej rzecz biorąc, w algorytmie liczona jest

wykładnicza ruchoma średnia gradientu i kwadrat z gradientu oraz parametry β_1 i β_2 , które kontrolują zakres liczenia średnich.

W momencie pisania tej pracy algorytm Adam jest najczęściej rekomendowanym jako domyślna metoda szkolenia głębokich sieci neuronowych, dlatego poniżej zamieszczono jego dokładny opis:

1. Wybierz wartość początkową $\epsilon, \beta_1, \beta_2$ oraz ustaw początkowe wartości zmiennych momentu 1 i 2 stopnia $s = 0$ i $r = 0$, wartość kroku czasowego $t = 0$ i stałą σ używaną do stabilizacji numerycznej.
2. Wybierz minipakiet złożony z m przykładów ze zbioru szkoleniowego.
3. Oblicz estymację gradientu $g = \frac{1}{m} \nabla \sum_{i=1}^m L(x_i, y_i, f)$ i zwięksź t o 1
4. aktualizuj estymację pierwszego momentu. $s \mapsto \beta_1 s + (1-\beta_1)g$
5. aktualizuj estymację drugiego momentu. $r \mapsto \beta_2 r + (1-\beta_2)g \odot g$
6. skoryguj obciążenie momentu pierwszego rzędu $s = \frac{s}{1-\beta_1^t}$
7. skoryguj obciążenie momentu drugiego rzędu $r = \frac{r}{1-\beta_2^t}$
8. Zastosuj aktualizację wartości funkcji celu równą $-\epsilon \frac{s}{\sqrt{r+\sigma}}$.
9. Jeżeli kryterium stopu nie zostało spełnione wróć do kroku 2.

Ocena procesu szkolenia się sieci polega na obliczeniu odpowiednich miar i współczynników odzwierciedlających przybliżenie zbioru T albo W przez znalezione rozwiązanie. W powszechnym problemie klasyfikacji binarnej, występującym również w tej pracy, w kontekście oceny efektywności algorytmów można wyszczególnić następujące parametry:

- *Fałszywie pozytywna klasyfikacja* (*FP* od ang. *False Positive*) – liczba obserwacji zaklasyfikowanych jako pozytywne, a należących do klasy obserwacji negatywnych.
- *Fałszywie negatywna klasyfikacja* (*FN* od ang. *False Negative*) – liczba obserwacji zaklasyfikowanych jako fałszywie negatywne, a należących do klasy obserwacji pozytywnych.

- *Prawdziwie pozytywna klasyfikacja* (*TP* od ang. *True Positive*) – liczba wyników poprawnie zaklasyfikowanych jako pozytywne.
- *Prawdziwie negatywna klasyfikacja* (*TN* od ang. *True Negative*) – liczba wyników poprawnie zaklasyfikowanych jako negatywne,

oraz miary:

- *Dokładność klasyfikacji* (ang. *Accuracy*) – $ACC = \frac{TP+TN}{TP+TN+FP+FN}$.
- *Czułość klasyfikacji* (ang. *Sensitivity*) – $TPR = \frac{TP}{TP+FN}$.
- *Swoistość klasyfikacji* (ang. *Specificity*) – $TNR = \frac{TN}{TN+FP}$.
- *Precyzja klasyfikacji* (ang. *Precision*) – $PPV = \frac{TP}{TP+FP}$.
- *Nieprecyzyjność klasyfikacji* (ang. *Negative precision*) – $NPV = \frac{TN}{TN+FN}$.

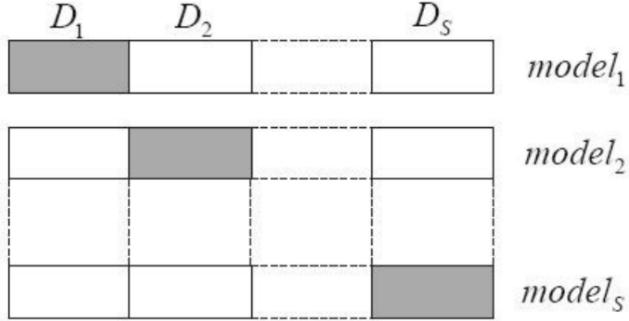
W problemach empirycznych dąży się do maksymalizowania TP, TN, ACC, TPR, TNR i PPV oraz minimalizowania FP, FN i NPV. Prawidłowe podejście polega również na wybraniu możliwie efektywnej architektury sieci. Problemy z tym związane zostaną omówione w kolejnych podsekcjach.

4.2.1 Problem nadmiernego dopasowania

Dążenie do najlepszego możliwego przybliżenia zbioru U wprowadza niepożądane zjawisko zwane *nadmiernym dopasowaniem*. Wtedy to dokładność klasyfikacji zbioru U jest wysoka lub nawet bezbłędna, natomiast znacznie niższa jest dokładność klasyfikacji zbioru testowego T i walidacyjnego W . W praktyce oznacza to, że model staje się mało użyteczny, gdyż wnioskowanie na nowych danych charakteryzuje się niską dokładnością.

Zatem ogólnym dążeniem w procesie uczenia się sieci jest osiągnięcie *maksymalnej generalizacji* klasyfikacji. Sieć o wysokim współczynniku generalizacji lepiej klasyfikuje ogólny zbiór wektorów wejściowych niż sieć, która ma niski współczynnik generalizacji i jest nadmiernie dopasowana do zbioru U . Właściwym działaniem jest więc ustalenie maksymalnie ogólnych, dostatecznych warunków poprawnej klasyfikacji, dzięki którym wzrasta prawdopodobieństwo, że przykład z poza zbioru U będzie poprawnie klasyfikowany.

W tym celu można zastosować *metodę oceny krzyżowej* (ang. *cross-validation*). Metoda ta polega na podziale zbioru uczącego na s segmentów D , z których każdy w innej iteracji służy jako zbiór testujący i walidacyjny, a pozostałe segmenty pełnią rolę zbioru uczącego. Podział zobrazowany jest na Rysunku 4.5. Stosując metodę oceny krzyżowej



Rysunek 4.5: Reprezentacja graficzna oceny krzyżowej.

dla różnych modeli sieci można stwierdzić, który z nich spełnia najlepiej kompromis między dobrą klasyfikacją zbioru U i wysoką generalizacją.

Kombinacja predykcji wielu różnych modeli jest bardzo wydajną metodą do poprawiania generalizacji i zmniejszenia błędu klasyfikacji na zbiorach innych niż U (zob. [23, 28]). Jednak współczesne sieci neuronowe, których przykłady zostały opisane w dalszych sekcjach mogą zawierać miliony parametrów i ich optymalizacja jest wymagająca obliczeniowo, dlatego w praktyce ogranicza się liczbę segmentów obierając $s \in \langle 5, 10 \rangle$.

Innym podejściem zaproponowanym w 2012 w [62] jest technika *dropout*, której główna idea bazuje na zerowaniu wyjścia neuronów sieci z prawdopodobieństwem 0,5 przy każdej iteracji treningu sieci. Neurony, które są w ten sposób tymczasowo dezaktywowane nie mają wpływu w danej iteracji na predykcję sieci i nie są uwzględniane przy wstecznej propagacji gradientu. Podejście to można porównać do treningu w każdej iteracji różnych modeli sieci. Dla przykładu w [73] wykazano, że metoda dropout wymaga jedynie 2 razy więcej iteracji do przybliżenia zbioru U , przy tym uzyskując znacznie lepszą generalizację.

Kluczowym składnikiem potrzebnym do treningu sieci i maksymalizacji generalizacji jest odpowiedni rozmiar zbioru danych. Jest to problem szeroko dyskutowany, gdyż zwłaszcza w danych medycznych istnieje szereg ograniczeń związanych z dostępem i akwizycją odpowiedniego materiału badawczego (np. ograniczenia prawne, związane z prywatnością lub z etyką). W przypadku, gdy zgromadzenie odpowiedniego zbioru

danych jest niemożliwe, pewnym rozwiązaniem problemu jest zastosowanie metod jego sztucznego powiększania (ang. *data augmentation*).

Dla obrazów stosuje się metody *afinicznych przekształceń* zgodne z definicją algebraiczną:

$$\mathbf{o} \mapsto \mathbf{A}\mathbf{o} + \mathbf{b}, \quad (4.9)$$

gdzie A jest macierzą przekształcenia liniowego, a b wektorem przesunięcia. Jako przykłady takich przekształceń dla dwuwymiarowych obrazów można wymienić:

- *rotacje* – obrót obrazu o kąt θ , gdzie:

$$A = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \quad (4.10)$$

- *odbicie lustrzane* – odwrócenie kolejności pikseli w każdym wierszu, gdzie:

$$A = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \quad (4.11)$$

- *skalowanie* – zmiana rozmiaru obrazu o S , gdzie:

$$A = \begin{bmatrix} S_x & 0 \\ 0 & S_y \end{bmatrix} \quad (4.12)$$

- *translacje* – przesunięcie punktów obrazu o wektor b .

Analogiczne równania istnieją dla obrazów trójwymiarowych (zob. [61]). W określonych przypadkach używane są również nieafiniczne przekształcenia. Np. w [104] wykorzystano nieliniowe deformacje do powiększenia zbioru 30 obrazów mikroskopowych przedstawiających macierz komórkową i uzyskano znaczco lepsze wyniki detekcji komórek niż istniejące w 2015 najlepsze rozwiązania.

W przypadku danych medycznych należy szczególnie zwrócić uwagę, aby powiększony zbiór zawierał dane przypominające w rzeczywistości występujące przypadki np. nieduże obroty występujące u pacjentów skanowanych rezonansem magnetycznym lub niewielkie skalowania rozmiaru kości widocznych w Tomografii Komputerowej. Szeroką dyskusję prowadzi się również na temat wykorzystania sztucznie generowanych zbiorów danych o czym więcej można przeczytać w [82, 119].

4.2.2 Problem redukcji wymiarowości

Duży rozmiar wektora cech prowadzi do problemu nazwanego *przekleństwem wymiarowości* (ang. *curse of dimensionality*). Określenie zostało po raz pierwszy sformułowane przez Richarda Bellmana w latach 50-tych XX wieku. Naukowiec ten podczas swojej pracy obserwował algorytmy doskonale działające w 3 wymiarach, a prezentujące znacząco gorsze wyniki w hiperprzestrzeni (zob. [24]).

Problem przekleństwa wymiarowości ma dwie główne przyczyny: (1) nie wszystkie cechy są jednakowo znaczące w kontekście rozróżnienia danych; (2) w miarę wzrostu rozmiaru przestrzeni cech, liczba obserwacji w zbiorze uczącym potrzebnych do wiarygodnego oszacowania funkcji wyjściowej rośnie wykładniczo.

Problem (1) jest szczególnie istotny w dość prostych algorytmach takich jak np. algorytm *K-najbliższych sąsiadów*, gdzie do poprawnego działania należy policzyć dystans pomiędzy sąsiednimi obserwacjami. Uwzględniając dużą liczbę nieistotnych cech jako argumenty funkcji dystansu uzyskuje się wyniki utrudniające lub nawet uniemożliwiające poprawną klasyfikację zbioru. Konieczna jest wówczas adaptacja wpływu poszczególnych cech np. poprzez wprowadzenie wektora wag.

Problem (2) ma duże implikacje w praktyce stosowania sieci neuronowych, gdyż posiadanie odpowiednio dużego, ustrukturyzowanego zbioru danych stanowi przeważnie wyzwanie. W poprzedniej podsekcji omówiono możliwość sztucznego powiększania zbioru danych. Inną opcją jest zmniejszenie rozmiaru wektora cech, co może być wykonane na dwa sposoby:

- wybór podzbioru istotnych cech o liczności $n' \ll n$,
- przekształcenie oryginalnych n zmiennych na nowy zbiór n' cech, gdzie ponownie $n' \ll n$.

W pierwszym przypadku, wybór podzbioru istotnych cech polega na określeniu minimalnego podzbioru, dla którego rozkład prawdopodobieństwa różnych klas obiektów jest jak najbliższy oryginalnemu rozkładowi uzyskanemu z wykorzystaniem wszystkich cech. Do tych zagadnień wykorzystywane są metody takie jak:

- *miary siły związku* – określające podobieństwo między rozkładami zmiennych losowych. Najczęściej stosowana jest *korelacja Pearsona*, której współczynnik r

dla dwóch zmiennych losowych X i Y zapisywany jest następującym wzorem:

$$r_{XY} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}, \quad (4.13)$$

gdzie X_i, Y_i to wartości kolejnych obserwacji, a \bar{X} i \bar{Y} to ich średnie. Innymi słowy tak zapisany współczynnik r_{XY} jest ilorazem *kowariancji* i iloczynu *odchylen standardowych* zmiennych X i Y .

- *miary entropii względnej* – określające rozbieżności między rozkładami zmiennych losowych. Najczęściej stosowana jest *dywergencja Kullbacka-Leiblera*, której współczynnik d_{KL} dla dwóch rozkładów prawdopodobieństwa p i q zapisany jest wzorem:

$$d_{KL}(p, q) = \sum_i p(i) \log_2 \frac{p(i)}{q(i)} \quad (4.14)$$

- *teoria zbiorów przybliżonych* – wykorzystująca informacje o elementach zbioru i klasyczną teorię zbiorów do porównywania rozkładów (zob. [71]).

Podobieństwo rozkładów ocenione na podstawie powyższych miar daje możliwość skutecznej redukcji wymiarowości z zachowaniem efektywności algorytmu.

W celu selekcji cech można również stosować różnego rodzaju regresje, specjalnie zaprojektowane do tego celu. Do najbardziej powszechnie stosowanych należą algorytmy regresji grzbietowej, LASSO (od ang. *Least Absolute Shrinkage and Selection Operator*) oraz Elastic Net. Różnią się one definicją zadania minimalizacji wektora cech w :

$$\text{Regresja grzbietowa; } \min_w \|X * w - Y\|_2^2 + \alpha \|w\|_2^2 \quad (4.15)$$

$$\text{LASSO; } \min_w \|X * w - Y\|_2^2 + \alpha \|w\|_1 \quad (4.16)$$

$$\text{Elastic Net; } \min_w \frac{1}{2n_{samples}} \|Xw - y\|_2^2 + \alpha\rho\|w\|_1 + \frac{\alpha(1-\rho)}{2} \|w\|_2^2 \quad (4.17)$$

W szczególności istotny jest czynnik regulujący dołączony do klasycznego zadania regresji liniowej będącym pierwszym elementem każdego zadania minimalizacji. W regresji grzbietowej dodatkowo dodawany jest warunek aby norma L2 wektora w była minimalizowana, dzięki czemu wymuszane jest aby poszczególne współczynniki nie były za duże. W regresji LASSO stosowana jest norma L1, dzięki czemu preferowane są rozwiązania, w których wektor jest rzadki, co przekłada się na efektywny wybór cech z

grup silnie skorelowanych. Regresja Elastic Net łączy w sobie cechy regresji grzbietowej oraz LASSO, poprzez zastosowanie dwóch typów norm jednocześnie, których istotność względem siebie kontroluje parametr ρ . Minusem tego podejścia jest większa złożoność obliczeniowa i konieczność doboru wartości dodatkowego parametru. Z uwagi na definicję zadania minimalizacji, powyższe algorytmy znajdują zastosowania głównie w problemach, gdzie występują zależności liniowe. Do nieliniowych problemów stosuje się częściej bardziej skomplikowane algorytmy typu lasy losowe.

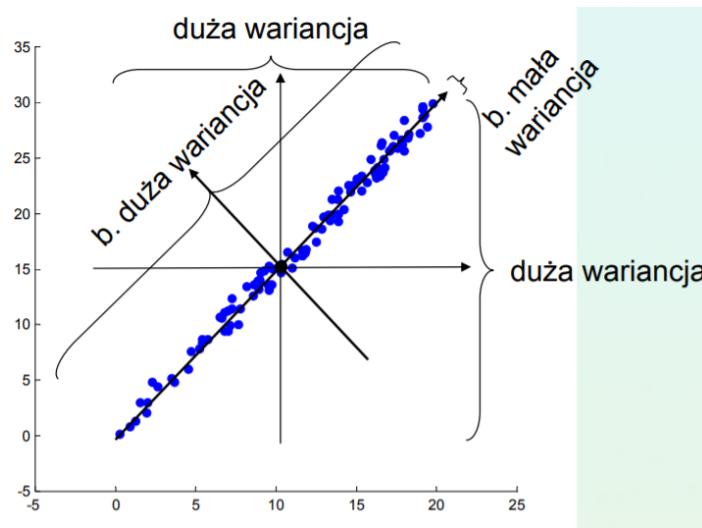
W kolejnym omawianym przypadku, tj. gdy zaistnieje potrzeba przekształcenia przestrzeni cech i wyliczenia nowego zbioru, metodą *state-of-the-art* jest algorytm *analizy składowych głównych* (ang. *Principal Component Analysis*, w skr. *PCA*) opracowany przez Karla Pearsona w 1901 r (zob. [99]).

Istotą PCA jest przekształcenie początkowych, skorelowanych cech w nowy zbiór nieskorelowanych zmiennych. Nowe zmienne, tzw. *składowe główne*, powstają z przekształcenia oryginalnych zmiennych skorelowanych, w taki sposób aby w maksymalnym stopniu wyjaśniać całkowitą wariancję w próbie cech oryginalnych. Wariancje składowych głównych są *wartościami własnymi* macierzy kowariancji oryginalnych zmiennych. Innymi słowy są to *wartości skalarne* (pojedyncze liczby) opisujące poziom wariancji w danych w stosunku do zadanego *ortogonalnych wektorów*, tj. gdzie iloczyn skalarny równy jest 0, co w przestrzeni Euklidesowej oznacza, że kat pomiędzy wektorami równy jest 90° .

Dla przykładu pierwsza składowa główna redukuje największą część zróżnicowania, druga kolejną, której nie redukowała poprzednia itd. Składowych głównych jest zawsze tyle ile wymiarów danych. Przykład dwuwymiarowy został przedstawiony na Rys. 4.6. Można zaobserwować, że nowa przestrzeń charakteryzuje się znacznie większym zróżnicowaniem poziomu wariancji niż przestrzeń oryginalna. Wybierając zatem tylko pierwszą składową uzyskana zostanie reprezentacja *skompromowana stratnie* (tzn. taka która nie daje gwarancji odtworzenia oryginalnych wartości), ale za to zachowująca znaczną część wariancji.

Opis szczegółowej procedury PCA wygląda następująco:

1. Oblicz macierz kowariancji: $S_x = X^T X$, gdzie X to macierz danych zawierająca obserwacje w wierszach. Macierz S_x jest symetryczna i pozwala ocenić: (1) wariancje zmiennych poprzez analizę elementów na głównej przekątnej; (2) zależności pomiędzy zmiennymi poprzez analizę elementów poza główną przekątną.



Rysunek 4.6: Dwuwymiarowa przestrzeń cech wraz z zaznaczonymi poziomami wariancji i składowymi głównymi.

2. Dokonaj diagonalizacji: $S_x = K L K^{-1}$, gdzie L to macierz diagonalna, a K to macierz odwracalna składająca się z wektorów własnych odpowiadających kolejnym wartościom własnym.
3. Utwórz macierz nowych zmiennych wykonując operację: $Y = XK$

Oprócz PCA, w kontekście redukcji wymiarowości można również wyróżnić następujące algorytmy:

1. *Multidimensional scaling*, MDS – bazuje na wyliczeniu odpowiednio zdefiniowanego dystansu między wartościami cech w oryginalnej przestrzeni i próbie zachowania tego dystansu w zredukowanej wymiarowości (zob. [31]).
2. *T-distributed stochastic neighbor embedding*, t-SNE – stosowany do wizualizacji struktur cech w różnych skalach (zob. [132]).
3. *Isomap* – algorytm nieliniowej redukcji wymiarowości pozwalający na gwarantowane znalezienie globalnego minimum (zob. [131]).
4. *Independent Component Analysis*, ICA – stosowany do znalezienia reprezentacji danych składających się z niezależnych elementów (zob. [76]).
5. *Latent Semantic Analysis*, LSA – stosowany głównie w problemach dotyczących przetwarzania języka naturalnego, zwłaszcza do znalezienia kontekstu użycia danego słowa poprzez analizę statystyczną dużych bloków tekstów (zob. [138]).

6. *Self Organizing Maps*, SOM – bazuje na zachowaniu własności topologicznych przestrzeni cech oryginalnych w zredukowanej wymiarowości (zob. [136]).

Zarówno przekleństwo wymiarowości, jak i wcześniej opisany problem nadmiernego dopasowania są zjawiskami bardzo często występującymi w praktycznych zastosowaniach algorytmów głębokiego uczenia się. Nie wyczerpują one jednak tematu. Szereg kolejnych wyzwań i rozwiązań został opisany w kolejnej sekcji przy okazji przedstawienia współczesnych topologii.

4.3 Przykłady współczesnych topologii

Współczesne sieci konwolucyjne służące do klasyfikacji obrazów, podobnie jak u protoplasty tj. architektury LeNet, składają się z dwóch głównych części: ekstraktora cech i klasyfikatora. W porównaniu jednak do pierwszych topologii, różnorodność warstw uległa zwiększeniu. Do najczęściej stosowanych komponentów można zaliczyć:

- *warstwy konwolucyjne* – grupują filtry używane do ekstrakcji cech obrazowych różnego poziomu (np. krawędzie, skupiska, obiekty);
- *warstwy max-pool* – stosowane do nieliniowej redukcji wymiarowości. Z obszaru $n \times n$ w obrazie wyliczana jest największa wartość ² $I(x, y)$. W praktyce zazwyczaj $n=2$, podobnie jak wartość *kroku* tj. odstępu pomiędzy kolejnymi próbkowaniami funkcji I .
- *warstwy aktywacji* – zawierające funkcje aktywacji neuronów. We współczesnych architekturach zamiast funkcji sigmoidalnych, czy progowych stosuje się najczęściej funkcję ReLU:

$$f(x) = \max(0, x). \quad (4.18)$$

Jest ona mniej wymagająca obliczeniowo niż funkcja sigmoid, a dla współczesnych topologii o dużej liczbie parametrów testy empiryczne wykazały, że równie dobrze wpływa na uzyskiwane wyniki dokładności (zob. [73]).

- *warstwy fully connected*, FC – składające się z neuronów, z których każdy jest połączony z każdym neuronem kolejnej warstwy. Występują zazwyczaj w ostatniej

²W uzasadnionych przypadkach stosuje się również obliczanie średniej, normy $L2 = \|x\| = \sqrt{\sum_{k=1}^n |x_k|^2}$ i innych współczynników.

części sieci konwolucyjnych tj. w klasyfikatorze.

- warstwy normalizacji³ – normalizujące wyjście poprzedzającej warstwy do pewnego określonego przedziału, zazwyczaj (-1,1). Przykładem jest operacja *Local Response Normalization* dana wzorem:

$$b_{x,y}^i = a_{x,y}^i / \left(k + \alpha \sum_{j=\max(0,i-n/2)}^{\min(N-1,i+n/2)} (a_{x,y}^j)^2 \right)^\beta, \quad (4.19)$$

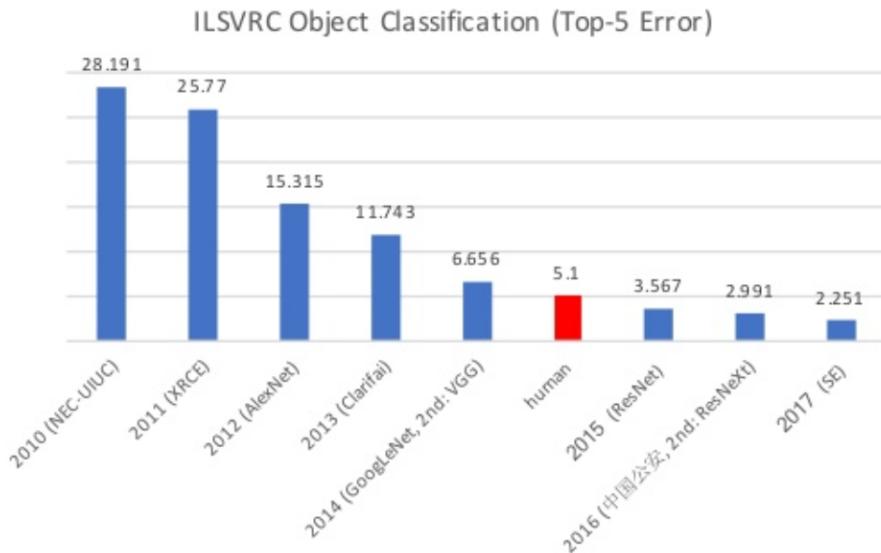
gdzie $a_{x,y}^i$ jest aktywacją danego neuronu, a k , α , n i β to stałe dobierane empirycznie na podstawie zbioru walidacyjnego (za [73]). Warstwy normalizacji wykorzystywane są w celu zrównoważenia poziomu wpływu poszczególnych neuronów na wynik końcowy.

Wprowadzone w ostatnich latach dedykowanych dla głębokiego uczenia się rozwiązań zarówno w warstwie sprzętowej jak i oprogramowania umożliwiło tworzenie architektur o wysokim stopniu komplikacji. W głównej mierze rozwiązania te umożliwiły optymalizację fazy szkolenia omówionej w poprzedniej sekcji jak również *fazy wnioskowania*, gdzie wytrenowana sieć użyta jest do przetwarzania kolejnych obserwacji. W warstwie sprzętowej wyróżnić można dedykowane akceleratory do operacji macierzowych takie jak [89], stworzone przez firmę NVIDIA *Tensor Processing Unit* (w skr. *TPU*) czy opracowane przez firmę Intel procesory Nervana [74].

W warstwie oprogramowania należy wspomnieć o rozwiązaniach takich jak TensorRT [11], które minimalizują czasy przetwarzania sygnału przez wytrenowaną sieć np. stosując optymalizację reprezentacji liczb zmiennoprzecinkowych lub przetwarzanych macierzy. Dynamicznie rozwijają się również framework do zoptymalizowanych obliczeń z udziałem głębokich sieci neuronowych np.: Caffe, Caffe2, TensorFlow, Theano, PyTorch czy MXNet. Ich porównanie osadzone w kontekście aplikacji medycznych można znaleźć w [46].

Postęp w rozwoju współczesnych sieci konwolucyjnych doskonale odzwierciedla progres w rezultatach konkursu ImageNet Large Scale Visual Recognition Competition (w skr. ILSVRC) przedstawiony na Rys. 4.7. Prezentowane wyniki dotyczą wartości błędu klasyfikacji na zbiorze danych [40] o nazwie ImageNet, uzyskiwanego w kolejnych latach przez zwycięskie algorytmy biorące udział w konkursie. *Błąd top-n* należy rozumieć jako zdarzenie, w którym dla danego obrazka w n wskazanych przez algorytm

³Coraz rzadziej stosowane w praktyce z uwagi na znikomy wpływ na rezultaty.



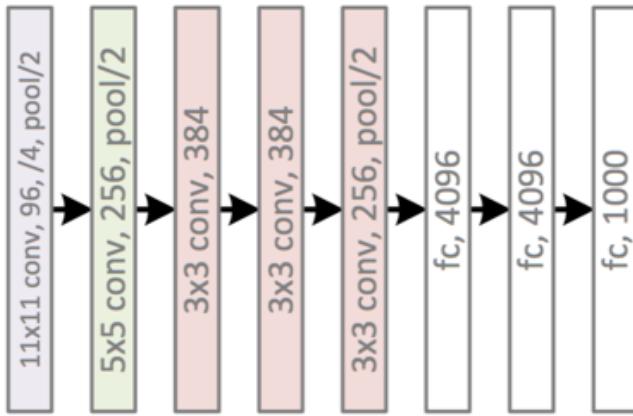
Rysunek 4.7: Błąd top-5 klasyfikacji obiektów w kolejnych latach uzyskiwany przez zwycięzców konkursu ILSVRC

najbardziej prawdopodobnych etykietach nie było poprawnej. W zakresie zmniejszenia wartości błędu top- n , znaczący progres dokonał się w 2012 roku, gdzie błąd top-5 zmalał o 10.4 punktów procentowych, co było rezultatem działania nowej architektury sieci konwolucyjnej nazwanej AlexNet. W kolejnych latach sieci konwolucyjne deklasowały inne podejścia doprowadzając w 2015 roku do spadku błędu top-5 do poziomu 3,5%, co jest uznawane za poziom lepszy niż możliwości ludzkiej klasyfikacji zbioru ImageNet. Lata 2016 i 2017 to intensywne prace nad synergią i złożeniami różnego rodzaju modeli, które w konsekwencji doprowadziły do obniżenia wartości błędu top-5 do poziomu 2,2%. W kolejnych podsekcjach zostanie dokładniej omówiona ewolucja zwycięskich architektur z konkursu ILSVRC.

4.3.1 AlexNet

Sieć AlexNet, której nazwa pochodzi od imienia głównego twórcy tej architektury Alexa Krizhevsky, zawiera blisko 60 milionów parametrów i 650 tysięcy neuronów. Architekturę zaprezentowano na Rys. 4.8

W skład topologii wchodzi pięć warstw konwolucyjnych i trzy typu fully-connected. Po pierwszej, drugiej i piątej warstwie konwolucyjnej występują operacje typu max-pool



Rysunek 4.8: Schemat architektury AlexNet.

z maską o wymiarach 2×2 ⁴.

Pierwsza warstwa konwolucyjna przyjmuje na wejściu dane o wymiarze $227 \times 227 \times 3$ ⁵, na których wykonywana jest operacja splotu z 96 filtrami z maską o wymiarach $11 \times 11 \times 3$ i krokiem 4. W rezultacie (uwzględniając również operację max-pool) objętość wynikowa przekazywana do kolejnej warstwy ma wymiar $27 \times 27 \times 96$. W drugiej warstwie konwolucyjnej wykonywana jest operacja splotu z 256 filtrami z maską o wymiarach $5 \times 5 \times 96$. Wymiar objętości wynikowej zostaje ponownie zredukowany poprzez operacje max-pool do $13 \times 13 \times 256$. Kolejne 3 warstwy konwolucyjne są połączone bezpośrednio ze sobą. Trzecia warstwa zawiera 384 filtry z maską o wymiarze $3 \times 3 \times 256$, w skład czwartej wchodzi 384 filtry z maską o wymiarze $3 \times 3 \times 384$, a w piątej znajduje się 256 filtrów również z maską o wymiarze $3 \times 3 \times 384$. Końcowe dwie warstwy typu FC zawierają po 4096 neuronów, a ostatnia tyle neuronów ile klas występuje w ostatecznym podziale – w oryginalnej pracy było to 1000 (por. [73]).

W celu lepszego zrozumienia przetwarzania sygnału wejściowego przez sieć poniżej przedstawiono przykład algorytmu wykorzystywanego dla pierwszej warstwy konwolucyjnej opisywanej topologii:

- Z danych wejściowych o wymiarze $[227 \times 227 \times 3]$ wybierany jest co czwarty blok o wymiarach $[11 \times 11 \times 3]$ (zarówno wzdłuż wysokości jak i szerokości). W rezultacie, nie uwzględniając krawędzi obrazu, otrzymywanych jest 217 punktów w każdym rzędzie i w kolumnie, w których mieści się $[55 \times 55]$ tj. 3025 bloków.

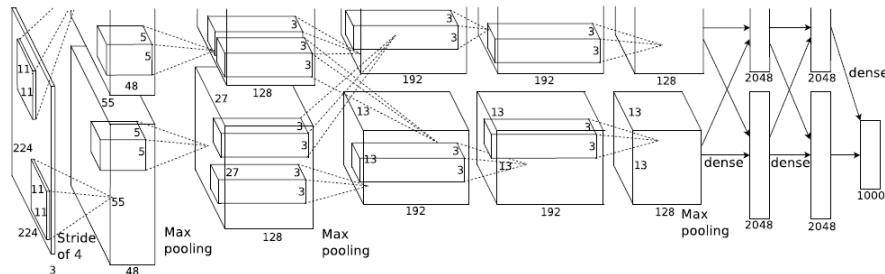
⁴Autorzy pracy podają też przykłady użycia masek o wymiarze 2×3 , które nakładają się w przestrzeni funkcji obrazowej. Nie znalazły one jednak miejsca w finalnej implementacji.

⁵gdzie 3 jest liczbą kanałów kodujących kolor obrazka.

2. Zarówno $11 \times 11 \times 3 = 363$ wagi znajdujące się w 96 filtrach jak i wartości 363 punktów obrazowych znajdujących się 3025 blokach są przedstawiane w postaci macierzy A o wymiarach $[96 \times 363]$ i B o wymiarach $[363 \times 3025]$.
3. Następnie liczony jest iloczyn skalarny w postaci $A^T B = C$, gdzie nowa, wyjściowa macierz C ma wymiar $[96 \times 3025]$.
4. Rezultat w postaci macierzy C ponownie przewymiarowany jest na postać $[55 \times 55 \times 96]$.

W architekturze jako funkcję aktywacji neuronów wykorzystano ReLU, co znaczaco przyspieszyło trening sieci. Dla przykładu uzyskano 6-krotne przyspieszenie treningu dla danych [72], CIFAR-10 w stosunku do tej samej topologii wykorzystującej sigmoidalną funkcję aktywacji. Ponieważ funkcja ReLU nie posiada górnego ograniczenia neurony teoretycznie mogą posiadać nieograniczone wartości funkcji aktywacji. W celu polepszenia kontrastu pomiędzy neuronami i wydobycia tych, które na tle innych się wyróżniają, zastosowano normalizację zgodną ze wzorem 4.19. W wyniku czego uzyskano redukcję błędu klasyfikacji top-5 o wartość 1,2 punktu procentowego.

W kontekście zwiększenia efektywności treningu zastosowano również powiększenie rozmiaru danych poprzez rotacje i modyfikacje funkcji obrazowej z wykorzystaniem czynników głównych (zob. [73]) zmniejszając błąd top-1 o 1%. Zastosowano też technikę dropout opisaną w 4.2.1. Ostatecznie wprowadzono także trening z wykorzystaniem wielu GPU (zob. Rys. 4.9).



Rysunek 4.9: Topologia architektury AlexNet z podziałem na dwa akceleratory GPU.

Topologia z podziałem na 2 karty zwiększyła dwukrotnie sumaryczną pamięć i pozwoliła na kolokację parametrów sieci.

Praca Alexa Krizhevsky, Ilya Sutskever i Geoffrey'a Hinton zapoczątkowała wzrost zainteresowania technikami głębokiego uczenia się, co doprowadziło do publikacji kolejnych podobnych architektur. Do najbardziej znanych należą ZFNet z 2013 roku [143],

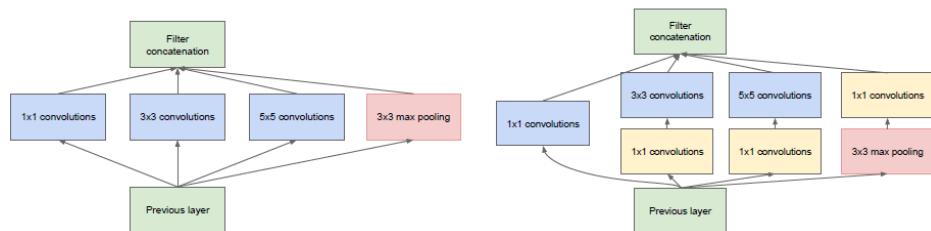
gdzie m.in. zastosowano zmniejszenie wymiaru maski stosowanego w filtrach pierwszej warstwy konwolucyjnej do 7×7 oraz VGGNet [118] z 2014 roku, gdzie zastosowano większą liczbę warstw konwolucyjnych z mniejszym wymiarem maski. Również w 2014 roku zaprezentowano innowacyjną koncepcję modułów sieci konwolucyjnych, co doprowadziło do zwycięstwa w ILSVRC. Idea ta została dokładniej opisana w kolejnej podsekcji.

4.3.2 GoogLeNet

Architekturę o nazwie GoogLeNet zaprezentowano w 2014 r. w pracy [126]. Nazwa architektury pochodzi od nazwy zwycięskiego zespołu startującego w ILSVRC 2014, składającego się z pracowników firmy Google. Oryginalnie topologia składała się z 22 warstw i zawierała około 5 mln parametrów (12 razy mniej niż w przypadku sieci AlexNet).

Redukcję liczby parametrów przy jednoczesnym podwyższeniu dokładności klasyfikacji obiektów udało się uzyskać poprzez poszukiwania konstrukcji optymalnych lokalnych topologii i ich połączeń. Mianowicie, wiadomo że duża część funkcji aktywacji neuronów przyjmuje wartość 0 lub jest redundantna z powodu wysokiej korelacji między sobą (zob. [20]). Matematyka dotycząca przetwarzania *macierzy rzadkich*, tj. gdzie przeważająca liczba elementów przyjmuje wartość 0, jest dobrze znana (zob. np. [93]). Jednak implementacje bibliotek do obliczeń związanych z algebrą liniową są zoptymalizowane pod kątem *macierzy gęstych*, gdzie przeważająca liczba elementów przyjmuje wartości różne od 0 (zob. [73, 120]).

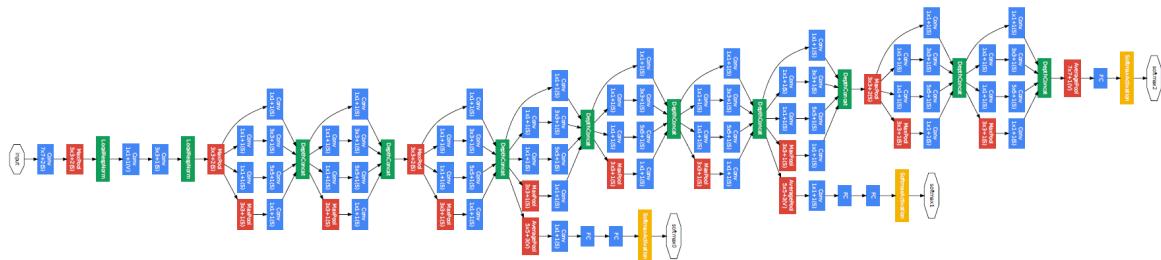
Ideą modułu incepcji zaproponowanego przez twórców GoogLeNet jest aproksymacja rzadkich macierzy z użyciem komponentów o gęstej strukturze. Takie komponenty nazywano *modułami incepcji* (ang. *inception modules*), a ich przykłady pokazano na Rys. 4.10.



Rysunek 4.10: Topologia architektury AlexNet z podziałem na dwa akceleratory GPU.

(a) przedstawia naiwną formę modułu incepcji, gdzie grupowane są operacje filtrów z maską o wymiarach 5×5 , 3×3 , 1×1 oraz operacja max-pool. (b) prezentuje koncepcję zoptymalizowaną obliczeniowo gdzie filtry z maską 1×1 służą do redukcji wymiarowości i używane są bezpośrednio przed splotami z bardziej wymagającymi obliczeniowo splotami 5×5 i 3×3 .

Przy pomocy złożenia różnego rodzaju modułów incepcji otrzymano topologię zaprezentowaną na Rys. 4.11



Rysunek 4.11: Topologia architektury GoogleNet

Dokładne zestawienie parametrów znajduje się w Tabeli 4.1

type	patch size/ stride	output size	depth	# 1×1	# 3×3 reduce	# 3×3	# 5×5 reduce	# 5×5	pool proj	params	ops
convolution	$7 \times 7 / 2$	$112 \times 112 \times 64$	1							2.7K	34M
max pool	$3 \times 3 / 2$	$56 \times 56 \times 64$	0								
convolution	$3 \times 3 / 1$	$56 \times 56 \times 192$	2		64	192				112K	360M
max pool	$3 \times 3 / 2$	$28 \times 28 \times 192$	0								
inception (3a)		$28 \times 28 \times 256$	2	64	96	128	16	32	32	159K	128M
inception (3b)		$28 \times 28 \times 480$	2	128	128	192	32	96	64	380K	304M
max pool	$3 \times 3 / 2$	$14 \times 14 \times 480$	0								
inception (4a)		$14 \times 14 \times 512$	2	192	96	208	16	48	64	364K	73M
inception (4b)		$14 \times 14 \times 512$	2	160	112	224	24	64	64	437K	88M
inception (4c)		$14 \times 14 \times 512$	2	128	128	256	24	64	64	463K	100M
inception (4d)		$14 \times 14 \times 528$	2	112	144	288	32	64	64	580K	119M
inception (4e)		$14 \times 14 \times 832$	2	256	160	320	32	128	128	840K	170M
max pool	$3 \times 3 / 2$	$7 \times 7 \times 832$	0								
inception (5a)		$7 \times 7 \times 832$	2	256	160	320	32	128	128	1072K	54M
inception (5b)		$7 \times 7 \times 1024$	2	384	192	384	48	128	128	1388K	71M
avg pool	$7 \times 7 / 1$	$1 \times 1 \times 1024$	0								
dropout (40%)		$1 \times 1 \times 1024$	0								
linear		$1 \times 1 \times 1000$	1							1000K	1M
softmax		$1 \times 1 \times 1000$	0								

Tabela 4.1: Parametry architektury GoogleNet

Ważną cechą sieci GoogleNet jest brak warstw typu FC na zakończeniu, gdzie w przypadku sieci AlexNet znajduje się około 90% parametrów. Końcowe wnioskowanie jest realizowane na podstawie wartości średniej z dwuwymiarowych map cech.

Dla lepszego zrozumienia idei redukcji wymiarowości realizowanej przez moduły incepcji, podobnie jak w przypadku sieci AlexNet, przeanalizowane zostanie działanie pierwszego modułu w topologii z Rys. 4.11. Moduł zawiera 128 filtrów z maskami o wymiarach 3×3 i 32 filtry z maskami o wymiarach 5×5 . Dane na wejściu modułu mają 192 kanały (zob. Tabela 4.1). Dla przykładu, rząd wielkości obliczeń operacji splotów 32 filtrów 5×5 wynosi $25 \times 32 \times 192 = 153\,600$ i dalej wzrastałby z głębokością sieci. W celu zapobiegnięcia nadmiarowi obliczeń stosowana jest redukcja z użyciem 16 filtrów z maską o wymiarach 1×1 . W efekcie rząd wielkości obliczeń spada do $16 \times 192 + 25 \times 32 \times 16 = 15\,876$, co pozwala na dalsze budowanie wielowarstwowych struktur.

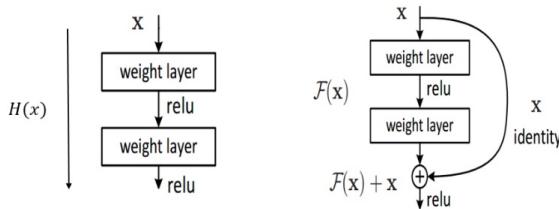
Topologia GoogLeNet jest wciąż rozwijana. Po pierwszej prezentacji pojawiły się kolejne modernizacje wprowadzające dodatkowe faktoryzacje modułów jak w Inception-v2 lub normalizacje wartości wynikowych poszczególnych warstw jak w Inception-v3. Obie sieci zostały przedstawione w [127]. Kolejny innowacyjny pomysł, bazujący na dodatkowych połączeniach między blokami, został wprowadzony w 2015 roku w sieci ResNet opisanej w kolejnej podsekcji.

4.3.3 ResNet

Jednym z najbardziej oczywistych pomysłów na polepszenie dokładności działania sieci neuronowych jest zwiększenie liczby warstw. Jednak wraz ze wzrostem liczby warstw, trening takich architektur z użyciem tradycyjnych metod gradientowych (takich jak algorytm wstępnej propagacji błędu) staje się mniej wydajny. Problem wynika z faktu, że zmiana wartości sygnału na wyjściu sieci w odpowiedzi na sygnał wejściowy jest mniejsza wraz ze wzrostem liczby warstw. W takiej sytuacji gradient wyliczany na podstawie sygnału będącego różnicą pomiędzy sygnałem wejściowym a wyjściowym może przyjmować wartości bliskie 0 uniemożliwiając dalszy postęp uczenia się. Problem zanikającego gradientu (ang. *vanishing gradient problem*) rozwiązywany jest poprzez zastosowanie normalizacji oraz nieliniowych funkcji aktywacji. Dzięki tym mechanizmom algorytm treningu głębokich sieci neuronowych w większej liczbie przypadków zbiega do użytecznego minimum lokalnego.

W momencie znalezienia takiego minimum dodanie kolejnych warstw i parametrów sieci jest redundantne, a nawet prowadzi do pogorszenia wyników treningu sieci. Zjawisko to nosi nazwę *degradacji treningu* (ang. *degradation problem*). Twórcy architektury ResNet, przedstawionej w [57], zaproponowali rozwiązanie tego problemu

poprzez implementację bloków rezydualnych (ang. *Residuum Units*) zawierających dodatkowe, skrótowe połączenia (ang. *skip connections*) pomiędzy wejściem a wyjściem bloków. Porównanie schematów funkcjonalnych nowych bloków i wcześniej istniejącego rozwiązania stosowanego np. w AlexNet zostało przedstawione na Rys. 4.12.



Rysunek 4.12: Schemat funkcjonalny pojedynczego bloku w architekturze ResNet.

Ogólną postać równania bloku rezydualnego można zapisać następująco:

$$\begin{aligned} y_l &= h(x_l) + F(x_l, W_l), \\ x_{l+1} &= f(y_l), \end{aligned} \tag{4.20}$$

gdzie x_l i x_{l+1} stanowią sygnał wejściowy i wyjściowy l -tego bloku. F stanowi funkcję rezydualną optymalizowaną podczas treningu sieci, $h(x_l)$ stanowi funkcję przekształcenia sygnału x_l przekazywanego skrótnym połączeniem, f jest funkcją ReLU, a W stanowi macierz wag.

Funkcja $h(x_l)$ jest funkcją tożsamościową, a zatem $h(x_l) = x_l$. Żeby uzasadnić ten wybór należy rozważyć propagację gradientu wewnątrz sieci składającej się z bloków rezydualnych. Dla każdego L -tego bloku zachodzi równanie:

$$x_L = x_l + \sum_{i=l}^{L-1} F(x_i, W_i) \tag{4.21}$$

Korzystając z reguły łańcuchowej można zapisać równanie na gradient funkcji kosztu ε :

$$\frac{\partial \varepsilon}{\partial x} = \frac{\partial \varepsilon}{\partial x_L} \frac{\partial x_L}{\partial x_l} = \frac{\partial \varepsilon}{\partial x_L} \left(1 + \frac{\partial}{\partial x_l} \sum_{i=l}^{L-1} F(x_i, W_i) \right) \tag{4.22}$$

z czego wynika, że gradient może być podzielony na dwie addytywne składowe: (1) $w = \frac{\partial \varepsilon}{\partial x_L}$ propagowaną bez wpływu na warstwy zawierające wagę i (2) $\lambda = \frac{\partial \varepsilon}{\partial x_L} \frac{\partial}{\partial x_l} \sum_{i=l}^{L-1} F(x_i, W_i)$ propagowaną przez nie.

Przykład propagacji gradientu w sieci składającej się z trzech bloków wygląda zatem

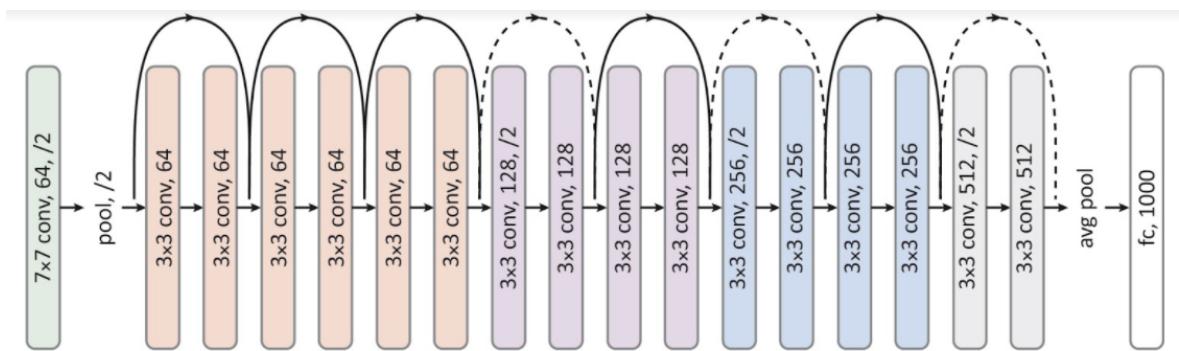
następującymi:

$$\frac{\partial \varepsilon}{\partial x_0} = \frac{\partial \varepsilon}{\partial x_3} * (w_2 + \lambda_2) * (w_1 + \lambda_1) * (w_0 + \lambda_0) \quad (4.23)$$

Wartości w są zazwyczaj znormalizowane do przedziału $(-1;1)$ można więc rozważyć 4 istotne przypadki równania 4.23:

1. $\lambda = 0$ – nie ma skrótowych połączeń, co odpowiada płaskiej strukturze sieci. Ponieważ wartości w są z przedziału $(-1;1)$ dodawanie kolejnych warstw wzmacnia wcześniej omówiony efekt zanikającego gradientu
2. $\lambda > 1$ – z każdą warstwą, sumaryczna wartość gradientu zwiększa się inkrementalni, co nazywane jest *problemem eksplozji gradientu* (ang. *exploding gradient problem*).
3. $\lambda < 1$ – przy założeniu, że $w + \lambda < 1$, dla sieci składających się z wielu warstw występuje problem zaniku gradientu, jak w przypadku 1. Natomiast, gdy $w + \lambda > 1$ podobnie jak w przypadku 2 może występować problem eksplozji gradientu
4. $\lambda = 1$ – wartości w są inkrementowane dokładnie o 1, co eliminuje problemy podane w przypadkach 1, 2 i 3 i stanowi uzasadnienie dla wyboru funkcji tożsamościowej $h(x_l)$ w architekturze ResNet.

Dokładny opis matematyczny funkcjonowania bloków rezydualnych wraz z dowodami znajduje się w [58]. Przykład topologii sieci składającej się z 8 bloków i łącznie 18 warstw tzw. ResNet-18, przedstawiono na Rys. 4.13.



Rysunek 4.13: Topologia architektury ResNet-18.

Pierwsza warstwa konwolucyjna zawiera filtry z maską o wymiarach 7×7 . W kolejnych zastosowano wymiar 3×3 . Zastosowanie mniejszych wymiarów masek niż w

AlexNet oraz podobnie jak w przypadku sieci GoogLeNet wyliczenie na końcu wartości średniej z dwuwymiarowych map cech zredukowało liczbę parametrów.

Architektura ResNet-18 jest najmniejszą z pojawiających się w literaturze przykładów tego typu. W praktyce, z powodzeniem wykorzystywano topologie składające się nawet z 1202 warstw (zob. [57]). W 2016 roku zaprezentowano w [125] hybrydę sieci GoogleNet i ResNet. Pracowano również nad bardziej złożonymi blokami, co w konsekwencji doprowadziło w 2017 roku do zaprezentowania architektury ResNetX w [141], która w wielu testach klasyfikacji różnych zbiorów okazała się być lepsza niż poprzednicy. Przegląd dotyczący historii tych prac można znaleźć w [12].

Sieć ResNet i jej warianty dla wielu testowych zbiorów danych takich jak ImageNet, CIFAR czy COCO [81] osiągnęły dokładność klasyfikacji porównywalną z możliwościami ludzkiego obserwatora. Dalszy progres był możliwy m.in. dzięki zastosowaniu synergii wielu modeli, co zostało opisane w kolejnej podsekcji.

4.3.4 Złożenia

Uczenie złożeń sieci (ang. *ensemble learning*) polega na wykorzystywaniu kilku modeli bazowych i wybranej metody ich synergii. W kontekście głębokiego uczenia się stosowane są różne metody kombinacji modeli bazowych (zob. [66]). Jako często stosowane przykłady można podać: uśrednianie, głosowanie, klasyfikacja Bayesa, generalizację stosów. Zostaną one kolejno omówione:

Uśrednianie

Uśrednianie jest prostą metodą kombinacji wyników predykcji. Najczęściej stosowane jest uśrednienie bez wag, gdzie suma wyników predykcji modeli bazowych podzielona jest przez ich liczbę. Uśredniać można bezpośrednio wyniki ostatecznej klasyfikacji jak również prawdopodobieństwa przynależności do odpowiednich klas, które są np. wynikiem *funkcji softmax*:

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}}, \quad (4.24)$$

używanej często bezpośrednio przed ostatnią warstwą sieci neuronowych dla z sygnałów wejściowych i j wyjściowych.

Główna zaletą uśredniania jest redukcja wariancji. Jest ona tym większa im bardziej nieskorelowane są wyniki predykcji modeli bazowych. Pomimo prostoty, tego rodzaju

koncepcja odnosiła już sukcesy m.in. w lasach losowych (zob. [28]).

Zastosowanie uśredniania przy silnie odstających od średniej najgorszych predykcjach znacząco obniża dokładność całego złożenia. Dlatego przy tak nieheterogenicznych modelach bazowych dających bardzo różne wyniki poszukiwane są inne metody.

Głosowanie

W głosowaniu stosuje się mechanizm zliczania przewidzianych przez modele bazowe etykiet. Etykieta, która została wybrana przez największą liczbę modeli bazowych jest obierana jako wynik ostatecznej predykcji. Jest to tzw. *głosowanie większościowe*.

W porównaniu do uśredniania, głosowanie jest mniej czułe na predykcje pojedynczych modeli. Wykorzystuje jednak jedynie informacje o przewidzianych etykietach, co utrudnia konstrukcję bardziej wyszukanych rozwiązań.

Klasyfikacja Bayesa

W przypadku tej metody, każdy model bazowy j postrzegany jest jako hipoteza h_j . Każda z hipotez posiada wagę proporcjonalną do prawdopodobieństwa zdarzenia, w którym dany zbiór trenujący zostałby wybrany z ogółu danych gdyby dana hipoteza była prawdziwa. Jest to tzw. optymalna klasyfikacja Bayesa, którą można zapisać następującym równaniem:

$$y = \operatorname{argmax}_{c_j \in C} \sum_{h_i \in H} P(c_j | h_i) P(T | h_i) P(h_i) \quad (4.25)$$

gdzie y to przewidziana etykieta, C jest zbiorem wszystkich możliwych klas, H to przestrzeń hipotez, a T to zbiór danych trenujących.

W praktyce z uwagi na dużą złożoność obliczeniową nie stosuje się optymalnej klasyfikacji Bayesa, a jedynie aproksymacje tej metody np.: BPA (od ang. *Bayesian parameter averaging*) [37], BMA (od ang. *Bayesian model averaging*) [108], czy też BMC (od ang. *Bayesian model combination*) [90].

Generalizacja stosów

Idea generalizacji stosów oryginalnie została zaproponowana w [139]. Wykorzystana została koncepcja *meta-uczenia*, a zatem konstrukcja nadzawanego klasyfikatora, który

rego zadaniem jest wybór optymalnego wektora wag a dla stosu s predykcji dla danych x :

$$s(x) = \sum_{i=1}^m a_i s_i(x) \quad (4.26)$$

W praktyce predykcje z modeli bazowych składowane są na stosie, a następnie klasyfikator nadzędny wykorzystuje je jako dane do treningu poprawnych wartości a wykorzystując jako odniesienie znane, poprawne etykiety.

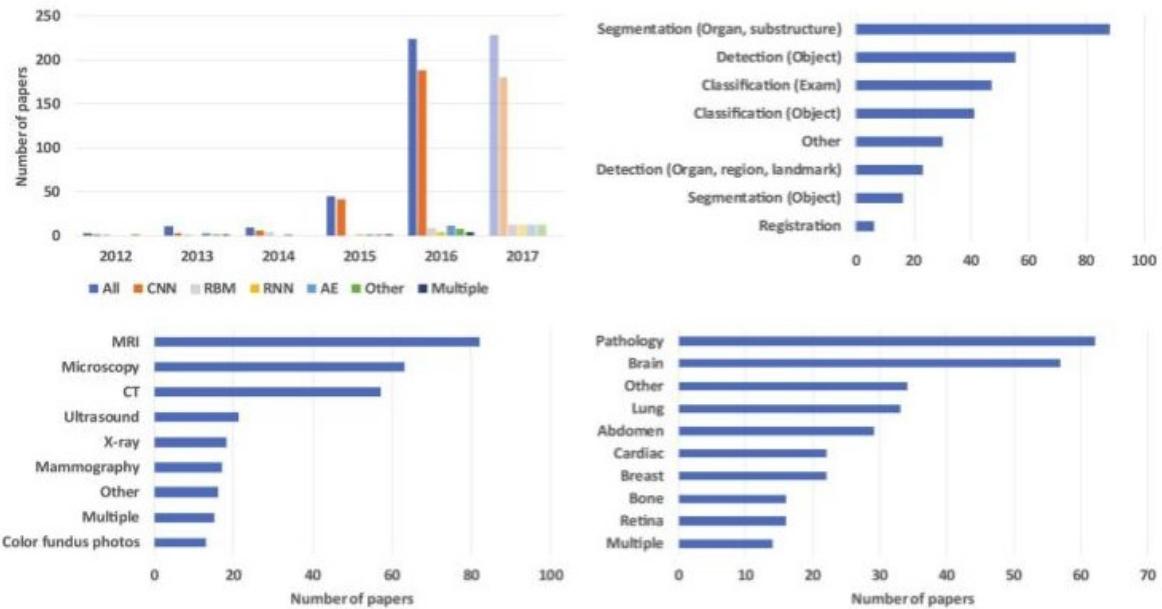
4.4 Zastosowania w medycynie

W 1994 roku ukazała się pierwsza praca, która w praktyce wykorzystywała mechanizmy związane z głębokim uczeniem się do przetwarzania obrazów medycznych (zob. [144]). Użyte wówczas sieci nazywano sieciami typu *shift-invariant*. Zastosowanie ich pozwoliło na eliminację 55% FP otrzymywanych przy wcześniejszych metodach stosowanych do detekcji skupisk mikro-zwapnień w mammografach. *Shift-invariant* oznaczało, że przesunięcie obrazu wejściowego nie powodowało zmian w klasyfikacji, co jest istotną wartością dodaną, z uwagi na specyfikę implementacji toru akwizycji danych w praktyce radiologicznej.

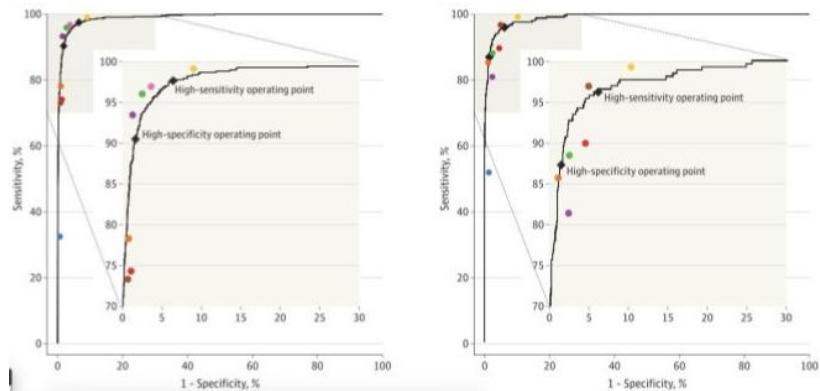
Po roku 2012 nastąpił znaczący wzrost zainteresowania metodami głębokiego uczenia się w medycynie. Obrazuje to praca [82] z 2017 roku, w której przytoczono statystyki medycznych publikacji zawierających słowa kluczowe związane z deep learning. Wybrane dane przedstawiono na Rys. 4.14.

Widoczny wzrost liczby publikacji nastąpił począwszy od 2015, co związane było z kilkuletnią adaptacją nowych metod w dziedzinie przetwarzania obrazów medycznych i gromadzeniem odpowiednich zbiorów danych. Lata 2016 i 2017 były pod pewnym względem przełomowe gdyż pojawiało się coraz więcej prac naukowych, w których przedstawiano rezultaty dokładności klasyfikacji na poziomie dorównującym ekspertom dziedzinowym.

Dla przykładu, w Listopadzie 2016 ukazała się praca [54] grupy Google Research z Mountain View w Kalifornii, gdzie zastosowano sieć GoogLeNet w wersji inception-v3 do zautomatyzowanej detekcji retinopatii cukrzycowej i cukrzycowego obrzęku plamki w obrazach dna oka. Wyniki porównano z panelem składającym się z 7 ekspertów, okulistów. Porównanie przedstawiono na Rys. 4.15.



Rysunek 4.14: Statystyki dotyczące publikacji medycznych zawierających słowa kluczowe związane z głębokim uczeniem się.

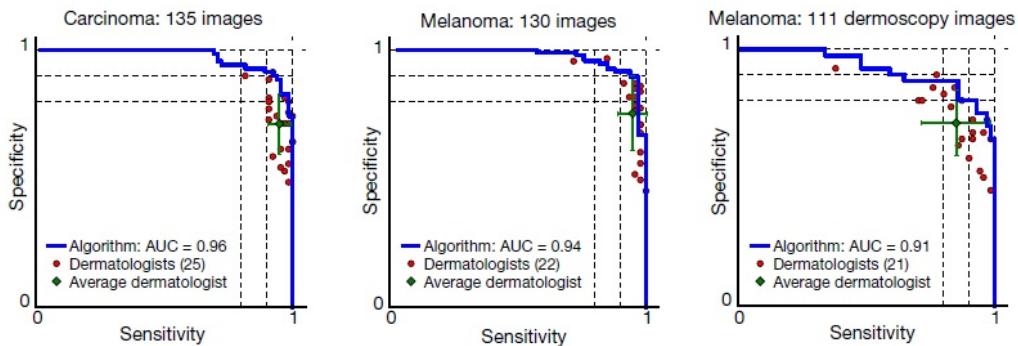


Rysunek 4.15: Porównanie automatycznej klasyfikacji retinopatii cukrzycowej i cukrzycowego obrzęku plamki z oceną panelu ekspertów.

Na wykresach dla dwóch zadań klasyfikacyjnych umieszczone krzywe reprezentujące zależność swoistości od czułości dla algorytmu automatycznego oraz 7 punktów oznaczających wynik oceny każdego z okulistów. Ogółem mniej niż połowa ekspertów uzyskała lepszy wynik niż algorytm sztucznej inteligencji.

Kolejna ciekawa praca tj. [48], pojawiła się w czasopiśmie Nature w styczniu 2017 roku i traktowała o automatycznej detekcji nowotworów skóry na podstawie zdjęć. Autorzy wykorzystali dane składające się z 129.450 obrazów klinicznych, na których

zobrazowano 2.032 różne schorzenia skóry. Ponownie do klasyfikacji wykorzystano sieć GoogleNet w wersji inception-v3. Wyniki klasyfikacji automatycznej porównano z oceną przeprowadzoną przez 21 certyfikowanych dermatologów. Przykład porównania zaprezentowano na Rys. 4.16.



Rysunek 4.16: Porównanie automatycznej klasyfikacji 3 chorób skóry z oceną ekspertów dermatologów.

Wykres przedstawia zależność czułości od swoistości. Czerwonymi punktami oznaczono wynik oceny poszczególnych ekspertów, a zielonym krzyżykiem wynik uśredniony. W każdym przypadku średnia ocena była gorsza od automatycznej klasyfikacji.

Obrazy medyczne nie są jedynymi danymi, które z powodzeniem są przetwarzane za pomocą metod głębokiego uczenia się. W lipcu 2017, przez grupę ze Stanford University została opublikowana praca [102] dotycząca klasyfikacji arytmii na podstawie szeregiów czasowych zapisanych na elektrokardiogramach. Autorzy wykorzystali dane z 64.121 elektrokardiogramów, próbkowanych z częstotliwością 200 Hz, pochodzących od 29.163 pacjentów. Zaprojektowano dedykowaną, 34-warstwową sieć konwolucyjną do detekcji 12 różnych dysfunkcji pracy serca, pracy prawidłowej i szumów (łącznie 14 klas). Wyniki klasyfikacji porównano z oceną prowadzoną przez 3 kardiologów. Średnia dokładność z oceny automatycznej wyniosła 80%, natomiast manualnej 72%.

Podobnych przykładów zostało opublikowanych dużo więcej. Architektura AlexNet z sukcesem była użyta do detekcji polipów w kolonoskopii w [129]. Sieć ResNet sprawdziła się w badaniach zrealizowanych w [47] w Mayo Clinic Rotschester. Dotyczyły one radio-genomiki i rozróżnienia zmian w mózgu bez konieczności biopsji. Złożenia natomiast z sukcesem zaaplikowano w pracach dotyczących detekcji nowotworów płuc, gdzie modele bazowe analizowały różne skale problemu (zob. [43]). W wielu pracach dotyczących radiologii raportuje się dokładność klasyfikacji automatycznej znacząco przewyższające możliwości dziedzinowych ekspertów np. [35, 51, 110, 137].

Powyższe przykłady pokazują, że dla szczególnych przypadków pewien element pracy eksperta zajmującego się danymi medycznymi (np. radiologa) może być z sukcesem wspomagany (lub nawet zastąpiony) przez algorytmy głębokiego uczenia się. Należy jednak podkreślić, że jest również szereg problemów wiążących się z wykorzystaniem tego rodzaju sztucznej inteligencji w medycynie. Do najważniejszych należą:

1. Gromadzenie dużych zbiorów danych z odpowiednimi etykietami.
2. Wykorzystanie heterogenicznych danych pochodzących np. z wielu urządzeń lub modalności.
3. Kalibracja i szacowanie niepewności wyników modeli.
4. Unifikacja modeli wykonujących podobne zadania.
5. Minimalizacja liczby parametrów modelu przy zachowaniu satysfakcjonującego poziomu dokładności.

Więcej na temat ograniczeń metod głębokiego uczenia się można przeczytać również w [88].

Dyskusja dotycząca tych problemów wciąż jest tematem wielu paneli dyskusyjnych i debat konferencyjnych (zob np. [6]). Najbardziej zaawansowane prace dotyczą problemu gromadzenia dużych zbiorów danych medycznych, co wymaga bliskiej współpracy ekspertów medycznych z ekspertami od uczenia maszynowego. Często konieczna jest również modyfikacja bądź tworzenie dedykowanych programów do akwizycji danych medycznych. Jako przykłady takich inicjatyw można wymienić programy Stanford Medicine [9], Harvard School of Medicine [33] czy Massachusetts General Hospital, które w swoich repozytoriach zgromadziły już dziesiątki milionów zdjęć radiologicznych (za [4]). Ponadto w roku 2018, na konferencji NVIDIA GTC (GPU Technology Conference) w San Jose (Kalifornia), Amerykańskie Stowarzyszenie Radiologii i stowarzyszenie MICCAI (od ang. *Medical Image Computing and Computer Assisted Intervention*) ogłosili porozumienie, co do wspólnej współpracy mającej na celu eliminacje barier legislacyjnych związanych ze współpracą przy pozyskiwaniu danych i wykorzystaniu algorytmów uczenia maszynowego.

Autor tej rozprawy jest świadom ograniczeń jakie są związane z wykorzystaniem algorytmów głębokiego uczenia się. Jednocześnie jednak nowe metody i ostatnie sukcesy zastosowań współczesnych sztucznych sieci neuronowych w aplikacjach medycznych stanowią silną motywację do przeprowadzenia własnych badań z ich wykorzystaniem.

Rozdział 5

Nowa metoda oceny procesu gojenia ścięgna Achillesa

W tym rozdziale zostanie zaprezentowana nowa metoda oceny procesu gojenia się ścięgna Achillesa bazująca na badaniach obrazowych rezonansu magnetycznego. W szczególności przedstawiony zostanie ilościowy opis, umożliwiający w zobietywizowany sposób ocenę morfologii tkanek widocznych w obrazach RM oraz nowatorskie podejście do automatycznego wyliczania wskazanych w opisie parametrów.

W chwili pisania tej pracy nie istnieje wedle najlepszej wiedzy autora podejście umożliwiające zobietywizowany i co ważniejsze automatyczny sposób oceny badań obrazowych prezentujących gojące się ścięgno Achillesa. Fakt ten implikuje trudności z integracją wyników radiologicznych z innymi formami oceny np. ze skalami testów funkcjonalnych takich jak ATRS, czy wynikami badań biomechanicznych. W rezultacie zmniejszona jest efektywność całego procesu oceny rehabilitacji.

Podczas próby rozwiązania wskazanego problemu autor tej pracy w szczególności skupił się na dwóch aspektach. Pierwszym z nich jest jakość generowanej automatycznie oceny, drugim czas akwizycji danych. W pierwszym przypadku jakość oceny automatycznej porównana została z oceną doświadczonego radiologa. W drugim, prace skupiły się wokół wyboru praktycznego protokołu, który zapewnił możliwie krótki udział pacjenta w badaniu. Celem jest poszukiwanie optimum, w którym maksymalizowana jest jakość oceny, a minimalizowany czas, zarówno radiologa jak i pacjenta, niezbędny do wykonania koniecznych czynności. Szczegóły realizacji tego celu przedstawiono w trzech kolejnych sekcjach rozpoczynając od przyjętej metodyki, następnie

opisując eksperymenty służące do doboru metod i ich parametryzacji, a kończąc na weryfikacji rozwiązania poprzez porównanie z oceną eksperta radiologa.

5.1 Metodyka

W tej sekcji została szczegółowo opisana proponowana metoda automatycznej oceny procesu gojenia się ścięgna Achillesa widocznego w badaniach obrazowych Rezonansu Magnetycznego. Dodatkowo scharakteryzowany został zbiór danych, który posłużył do opracowania rozwiązania, jak również ankieta walidacyjna stanowiąca wzorzec odniesienia dla przedstawionej metody.

Autor tej pracy, w proponowanym podejściu skorzystał z metod widzenia komputowego, a dokładniej z fuzji algorytmów sztucznej inteligencji i przetwarzania obrazów. W kontekście tej pracy, pierwsze znajdują swoje zastosowanie do ekstrakcji wektora cech dla danej reprezentacji obrazowej. Drugie, pozwalają uwzględnić wiedzę dziedzinową w procesie numerycznej oceny.

W przypadku algorytmów sztucznej inteligencji zastosowano opisane w Rozdziale 4 konwolucyjne sieci neuronowe, a dokładniej AlexNet, GoogLeNet (inceptionV3) i ResNet-18 (z osiemnastoma warstwami konwolucyjnymi). W pierwszej kolejności wykonano szkolenie podanych sieci dla problemu binarnej klasyfikacji tj. rozróżnienia obrazów chorego i zdrowego ścięgna. Następnie, część klasyfikująca została usunięta z topologii sieci, pozostawiając ekstraktor cech z parametrami zoptymalizowanymi pod kątem wydobycia istotnej informacji opisującej różnice między zdrową i chorą tkanką. Na tak otrzymanym wektorze, przeprowadzono redukcję wymiarowości z wykorzystaniem metody PCA (zob. 4.2.2). W wyniku przeprowadzonych eksperymentów, ostatecznie zdecydowano się uwzględnić 200 pierwszych czynników głównych w końcowym rozwiązaniu.

W przypadku metod przetwarzania obrazów zastosowano obliczenia cech z wydzielonego przez specjalistę radiologa ROI (od ang. *Region of Interest*) reprezentującego tkanki otoczone ościęgnem. Sumarycznie wyliczono 46 klasycznych cech obrazowych, w tym pole powierzchni ROI, 9 cech opisujących statystykę wartości pikseli i 36 cech Haralick'a opisujące teksturę (zob. [56]). W ramach statystyk scharakteryzowano: min., max., średnią, odchylenie standardowe, skośność, kurtozę, 25-percentyl, medianę oraz 75-percentyl. Natomiast w ramach cech Haralick'a: drugi moment kątowy, kontrast, korelację, wariancję, odwrotny moment różnicowy, sumę średnich, sumę wariancji, sumę

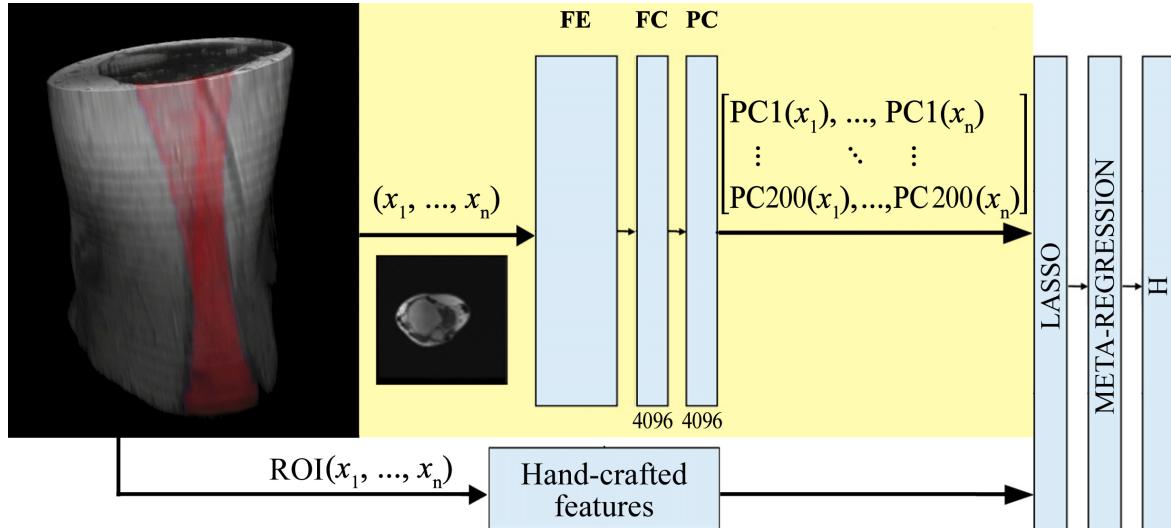
entropii, entropię, różnicę wariancji i maksimum prawdopodobieństwa. Cechy Haralick'a wyliczono dla 3 dystansów separacji $d=1, 5, 10$. Dokładna metoda prowadząca do selekcji powyższych cech została przedstawiona w pracy [97].

Na zbiorze wyliczonych 246 cech dokonano selekcji metodą LASSO (zob. pkt 4.2.2). W rezultacie dla wszystkich ocenianych parametrów uzyskano optymalne podzbiory o liczbowości poniżej 20 cech, ekstrahowane zarówno z wykorzystaniem sieci konwolucyjnych jak i klasycznych metod obrazowych. Predyktory te posłużyły do szkolenia algorytmu meta-regresji, przy użyciu którego wykonana została fuzja prowadząca do końcowej oceny badania realizowanego zaproponowaną przez autora tej pracy metryką H :

$$H = TM(R(x_1), R(x_2), \dots, R(x_n)), \quad (5.1)$$

gdzie TM jest średnią trymowaną z marginesami 2.5%, a $R(x_i)$ jest wynikiem regresji obliczonym dla przekroju osiowego x_i , gdzie i to indeks przekroju w trójwymiarowym badaniu RM.

Schemat proponowanej metody został zaprezentowany na Fig. 5.1. W ogólnym uję-



Rysunek 5.1: Schemat automatycznej metody oceny procesu gojenia się ścięgna Achillesa.

ciu, trójwymiarowe badanie RM dzielone jest na pojedyncze przekroje osiowe, które stanowią dane wejściowe do oznaczonej na żółto części opartej o metody głębokiego uczenia się. W tym komponencie dane przetwarzane są przez ekstraktor cech sieci AlexNet (FE), którego parametry wybrane w wyniku eksperymentów. Otrzymane cechy wstępnie grupowane są z wykorzystaniem warstwy gęstej (FC) i redukowane z 4096

pojedynczych wyjść aktywacyjnych do 200 czynników głównych (PC).

Równolegle, ROI z każdego z przekrojów stanowi dane wejściowe do obliczeń klasycznych cech obrazowych. Końcowa fuzja realizowana jest z wykorzystaniem metaregresji i metryki H . Takie podejście zapewnia numeryczny opis składający się z pojedynczej wartości dla całego badania RM.

5.1.1 Zbiór danych

Dane wykorzystane w tej pracy zostały zebrane w ramach projektu START "Wykorzystanie autologicznych mezenchymalnych komórek macierzystych w procesie regeneracji rekonstruowanego ścięgna Achillesa" finansowanego z konkursu STRATEGMED1 przez Narodowe Centrum Badań i Rozwoju. W ramach projektu, do badanej grupy zakwalifikowano 60 pacjentów po całkowitym zerwaniu ścięgna Achillesa i 29 ochotników stanowiących grupę odniesienia.

Kryteria wykluczające z grupy badanej zostały zdefiniowane tak, aby kwalifikowane osoby nie posiadały wcześniejszej historii urazów ścięgna Achillesa, ani przewlekłych chorób, które mogłyby spowodować degenerację tkanki ścięgnistej. Uwzględniono również warunki odbiegające od normy takie jak: otyłość, ciąża i infekcje. Ostatecznie, w ramach grupy pacjentów znalazło się 49 mężczyzn i 11 kobiet w wieku pomiędzy 24 a 49 lat, ze średnią 36 lat. Wszyscy pacjenci podpisali stosowne zgody wymagane do przeprowadzenia badań.

52-óch z pośród 60 pacjentów zerwało ścięgna podczas uprawiania sportu. Dokładniej były to: piłka nożna (19), squash/tenis (8), koszykówka (4), bieganie (3), podskoki (2), siatkówka (2) oraz badminton, box, piłka ręczna, fitness, cross-fit, skateboard, rugby i skakanka. Urazy w 31 przypadkach dotyczyły prawego ścięgna, a w 29 lewego. Średnia pozycja zerwania umiejscowiona była 57.11 mm ponad górnym konturem kości piętowej.

W okresie do tygodnia po urazie, pacjenci przeszli zabieg rekonstrukcji wykonany metodą szycia na otwartym ścięgnie (zob. 3.1.4). Następnie rozpoczęto rehabilitację trwającą 12 miesięcy, monitorowaną z wykorzystaniem Rezonansu Magnetycznego i Ultrasonografii. Wykonano również badania biomechaniczne w środkowym i końcowym okresie rehabilitacji, zgodnie z protokołem opisanym w sekcji 3.4.

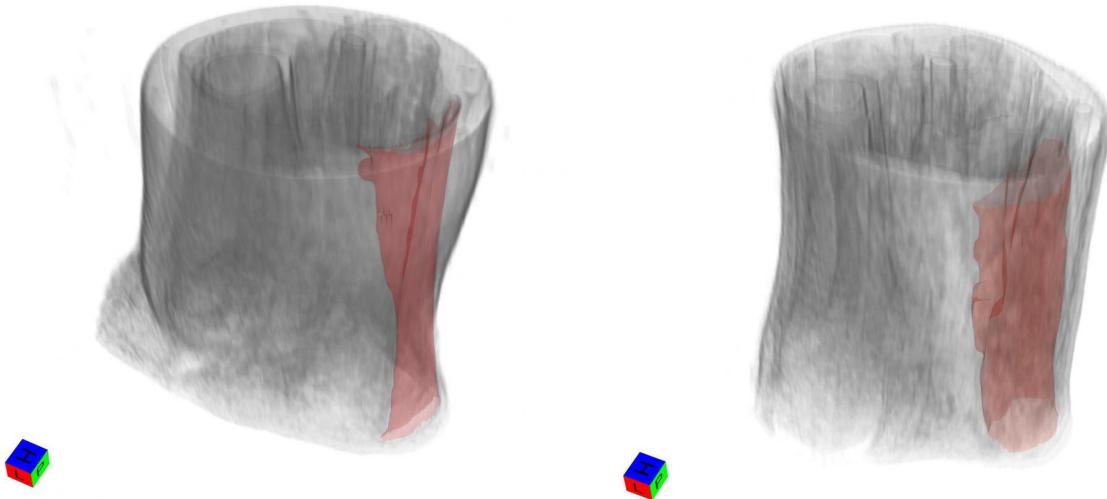
Przy akwizycji danych z wykorzystaniem RM posłużono się aparatem GE Signa HDxt 1.5T wyposażonym w cewkę Foot & Ankle dedykowaną do pomiarów w rejonie

dolnej kończyny. Każde z badań RM było wykonane z użyciem 7 sekwencji i łącznie 10 modalności opisanych szczegółowo w sekcji 3.2 tj.:

1. T1 zależne
2. T2 zależne
3. PD
4. T2 mapping
5. T2 * GRE
6. T2 * GRE TE_MIN
7. 3D FSPGR
 - In Phase Ideal
 - Out Phase Ideal
 - Fat Ideal
 - Water Ideal

Na grupie zdrowych ochotników przeprowadzono pojedyncze badanie, natomiast pacjentów skanowano 10-krotnie w odpowiednio zdefiniowanych odstępach czasowych. Pierwsze badanie odbyło się przed operacją, a następnych 9 po, odpowiednio w: 1, 3, 6, 9, 12, 20, 26, 40 i 52 tygodniu. W procesie zbierania danych wystąpiły problemy wynikające z braku obecności poszczególnych osób na badaniach lub niespójnością danych. Dlatego finalnie skompletowany zbiór składa się z homogenicznych badań 27 zdrowych ochotników (270 trójwymiarowych skanów RM, w skład których wchodzi wspomniane 10 modalności) oraz 59 badań pacjentów (5900 trójwymiarowych skanów RM z 10-oma modalnościami i 10-oma krokami czasowymi). Przykładowa wizualizacja danych pacjenta i osoby zdrowej znajduje się na Fig. 5.2. Obszar zajmowany przez tkankę ścięgnistą został oznaczony na czerwono. Należy zwrócić uwagę na różnicę w kształcie, która dla zdrowego przykładu jest zgodna z opisem anatomicznym ujętym w sekcji 3.1.1, natomiast dla przykładu chorego zawiera liczne deformacje.

Z pośród 59-iu zbadanych pacjentów, w ramach projektu START, udało się zgromadzić ustrukturyzowany opis radiologiczny dla 48 pacjentów (480 badań). Czterech z



Rysunek 5.2: Wizualizacja danych RM w sekwencji PD dla zdrowego (po lewej) i gojącego się (po prawej) ścięgna Achillesa.

nich zostało losowo wydzielonych na początku eksperymentów jako pacjenci testowi, a pozostały stworzyli grupę pacjentów treningowych.

Zbiory trójwymiarowe posłużyły do opracowania zbiorów dwuwymiarowych składających się z przekrojów. W ten sposób udało się m.in. zwiększyć liczbę próbek wejściowych potrzebnych do szkolenia sieci neuronowych. Dane RM charakteryzują się wysoką anizotropią i ich rozdzielcość wynosi przykładowo $512 \times 512 \times Z$, gdzie $Z \in (10 : 50)$ (wyjątkiem jest sekwencja 3D FSPGR charakteryzująca się niższą anizotropią, jednak większą ilością artefaktów wynikających z szybkiego czasu akwizycji). Zbiory oparte o RM zostały zatem opracowane w oparciu o przekroje osiowe 512×512 . Ważne jest również, że na każdym z tych przekrojów widoczne są elementy ścięgna Achillesa, co nie występuje w danych RM w przypadku przekrojów z pozostałych płaszczyzn.

Sumaryczna liczba tak uzyskanych próbek dwuwymiarowych wyniosła 11.725 (oznaczonych jako obrazy zdrowego ścięgna) i 138.604 (oznaczonych jako obrazy chorego ścięgna). Następnie, w zależności od eksperymentów służących do optymalizacji parametrów metody wykorzystano:

- powiększony zbiór treningowy RM – zawierający przekroje osiowe ze wszystkich badań pacjentów treningowych, realizowanych wszystkimi sekwencjami. Liczba przekrojów w tym zbiorze została powiększona poprzez zastosowanie transformacji odbicia lustrzanego i, w przypadku chorych przykładów, 10-ciu obrotów w zakre-

sie od -10 do 10 stopni. Finalnie zbiór zawiera 234.502 przekrojów oznaczonych jako zdrowych i 277.208 oznaczonych jako chorych.

- ograniczony zbiór treningowy RM – zawierający przekroje osiowe ze zbioru 44 pacjentów treningowych jedynie w sekwencji T2 * GRE TE_MIN (18.863 próbek).
- zbiór PCA – zawierający próbki z powiększonego zbioru treningowego ze wszystkich sekwencji RM dla 10 losowo wybranych pacjentów (44.813).

W celu wnioskowania i porównań rezultatów metody posłużono się danymi od 4-ech pacjentów testowych, z wykorzystaniem od 1 do 10 modalności RM. Szczegóły dot. poszczególnych wyborów znajdują się w opisach eksperymentów.

5.1.2 Wzorzec odniesienia

Do stworzenia wzorca odniesienia, w ramach projektu START, została powołana specjalna grupa robocza składająca się z eksperta radiologa, specjalistów ortopedów i ekspertów od komputerowo wspomaganej diagnostyki. Celem grupy było opracowanie ilościowego opisu, który odzwierciedla elementy, na podstawie których radiolog podejmuje subiektywną opinię. W wyniku prac stworzono zestaw sześciu parametrów oceniany w skali 0–7, opisujący proces gojenia się ścięgna Achillesa widoczny w obrazach RM:

1. Uszkodzenia śródścięgniste (SCT od ang. *Structural Changes within the Tendon*) – informuje o spójności włókien w obszarze ścięgna. Zawiera się w ROI. Ocena wykonywana zarówno w płaszczyźnie osiowej jak i strzałkowej. 0 – całkowity brak uszkodzeń, 1 – bardzo małe uszkodzenia, 2 – małe uszkodzenia, 3 – uszkodzenia średniej wielkości , 4 – dość duże uszkodzenia. 5 – duże uszkodzenia, 6 - bardzo duże uszkodzenia, 7 - ekstremalnie duże uszkodzenia.
2. Pogróżenie ścięgna (TT od ang. *Tendon Thickening* – najgrubsze miejsce w kierunku strzałkowym na widoczne na przekroju osiowym. Zawiera się w ROI. 0 – całkowity brak pogrubienia (3 mm-5 mm), 1 – bardzo małe pogrubienie - (6 mm), 2 – małe pogrubienie - (9 mm), 3 – średnie pogrubienie- (12 mm), 4 – dość duże pogrubienie -(15 mm), 5 duże pogrubienie - (18 mm), 6 – bardzo duże pogrubienie - (21 mm), 7 – ekstremalnie duże pogrubienie – (24 mm).

3. Ostrość granic ścięgna/rozgraniczenie (STE od ang. *Sharpness of the Tendon Edges*) – informuje o jakości granicy między tkankami ścięgnistymi i otoczeniem, w szczególności, czy brzeg jest fraktalny. Oceniane jest na zewnątrz ścięgna w płaszczyźnie osiowej. 0 – bardzo duża (idealna) ostrość, 1 – duża ostrość, 2 – dość duża ostrość, 3 – średnia ostrość, 4 – mała ostrość, 5 – bardzo mała ostrość, 6 – minimalna ostrość, 7 – całkowicie nieostre.
4. Obrzęk ścięgna (TE od ang. *Tendon Edema*) – informuje o anomalach w gromadzeniu się płynów w obszarze ścięgna. Zawiera się w ROI i jest oceniany na przekrojach osiowych. 0 – całkowity brak obrzęku, 1 – bardzo mały obrzęk, 2 – mały obrzęk, 3 – obrzęk średniej wielkości, 4 – dość duży obrzęk, 5 – duży obrzęk. 6 – bardzo duży obrzęk, 7 - ekstremalnie duży obrzęk.
5. Jednorodność ścięgna (TU od ang. *Tendon Uniformity*) – informuje o prawidłowej teksturze poszczególnych przekrojów osiowych jak i podobieństwu sąsiednich, mierzonym w płaszczyźnie strzałkowej. Zawiera się w ROI. 0 – całkowity brak niejednorodności (jednorodne), 1 – bardzo małe niejednorodności, 2 – małe niejednorodności, 3 – niejednorodności średniej wielkości, 4 – dość duże niejednorodności, 5 – duże niejednorodności, 6 – bardzo duże niejednorodności, 7 – ekstremalnie duże niejednorodności.
6. Obrzęk tkanek (TisE od ang. *Tissue Edema*) – informuje o powiększonym przedziale powięziowym ścięgna. 0 – całkowity brak obrzęku, 1 – bardzo mały obrzęk, 2 – mały obrzęk, 3 – obrzęk średniej wielkości, 4 – dość duży obrzęk, 5 – duży obrzęk. 6 – bardzo duży obrzęk, 7 - ekstremalnie duży obrzęk.

Przy użyciu ankiety zostały opisane obrazowania 3D dla 48 pacjentów.

5.2 Eksperymenty

W tej sekcji zaprezentowano eksperymenty mające na celu dobór poszczególnych komponentów i parametrów finalnej metody. W pierwszej kolejności zostało opisane badanie dotyczące szkolenia sieci neuronowych, celem wyboru ekstraktora cech, dającego możliwość rozróżnienia ścięgien zdrowych od tych po zerwaniu (chorych). Następnie zaprezentowano badanie dotyczące obliczenia wektora cech na wyjściu sieci neuronowej i redukcji jego wymiarowości. W kolejnej części przedstawiono wybór sekwencji

RM, stanowiących dane wejściowe do proponowanego w tej pracy algorytmu. Finalnie, w dwóch ostatnich podsekcjach opisano dobór optymalnego zestawu cech ekstrahowanych z użyciem sieci neuronowej i klasycznych metod przetwarzania obrazów oraz fuzję z wykorzystaniem meta-regresji.

5.2.1 Rozróżnienie ścięgna zdrowego i po zerwaniu

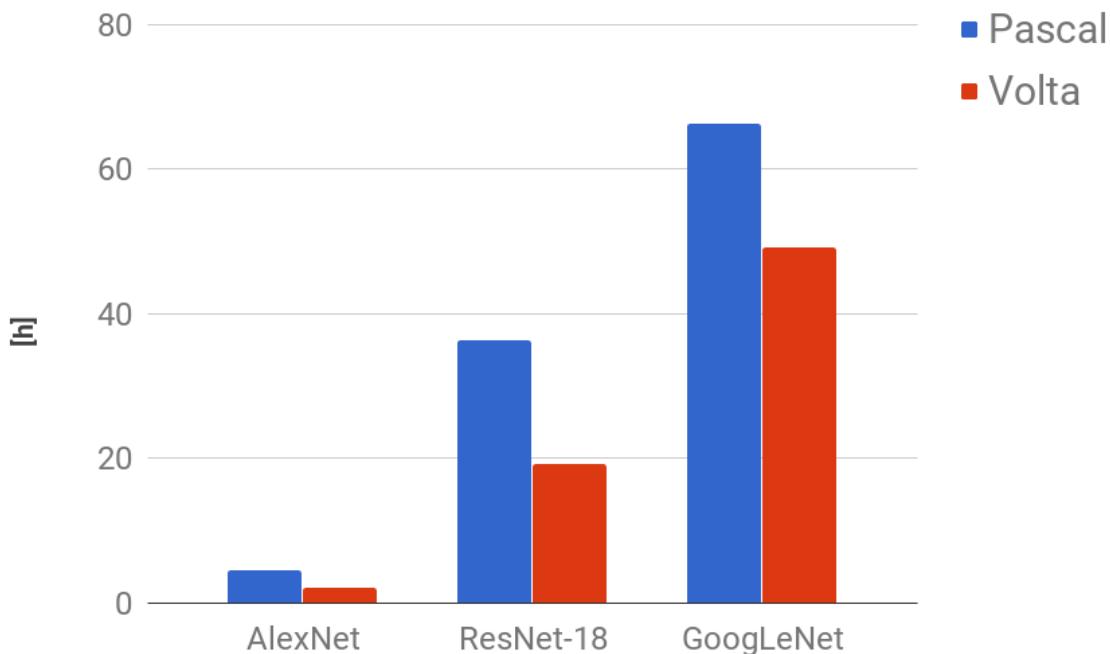
W tym badaniu porównane zostały współczesne architektury sieci neuronowych w zadaniu klasyfikacji binarnej, rozróżniającej obrazy zdrowego od chorego ścięgna. W porównaniu uwzględniono topologie AlexNet (zob. 4.3.1), GoogLeNet InceptionV3 (zob. 4.3.2) i ResNet-18 (zob. 4.3.3). W celu doboru hiperparametrów i optymalizacji parametrów sieci zastosowano szkolenie z wykorzystaniem kroswalidacji z podziałem na 5 segmentów (zob. 4.2.1). W każdym z 5 cykli, 3 segmenty były wykorzystane do treningu, 1 do walidacji i 1 do finalnych testów. Do celów tego badania wykorzystano powiększony zbiór treningowy RM. Wyniki zebrano w Tab. 5.1.

Tabela 5.1: Wyniki szkolenia sieci neuronowych z wykorzystaniem kroswalidacji z podziałem na 5 segmentów dla problemu klasyfikacji obrazów zdrowego i chorego ścięgna.

	Average	Min	Max	SD
AlexNet	99,19	99,15	99,24	0,04
ResNet-18	95,98	92,78	99,04	2,5
GoogleNet	99,83	99,68	99,91	0,1

Dla wszystkich architektur, wyniki dokładności klasyfikacji najlepszych modeli osiągnęły wartość powyżej 99%. Niemniej jednak, w przypadku architektury ResNet-18 najgorszy model klasyfikował z dokładnością 92,78%, co przełożyło się również na najgorszy średni wynik 95,98%. Fakt ten może wskazywać na tendencje do nadmiernego dopasowywania się architektury ResNet-18. Niższa generalizacja modelu może wynikać z faktu tendencji tej architektury do dużej wagi dla ekstrakcji cech szczegółowych przez jadra splotu o wymiarze 3×3 lub niedoboru liczby danych wejściowych w odniesieniu do liczby optymalizowanych parametrów.

W kolejnym kroku, dla wszystkich trzech architektur porównano również czasy szkolenia się dla zadanego problemu. Na Rys. 5.3 zestawiono łączny czas 5 iteracji procesu kroswalidacji. Porównano dwie architektury GPU, starszą Pascal P100 i nowszą Volta V100, zaprezentowaną w 2017 roku, charakteryzującą się udoskonalonym,



Rysunek 5.3: Porównanie czasów treningu architektur na dwóch architekturach GPU: Pascal i Volta.

osobnym układem optymalizowanym pod kątem operacji tensorowych realizowanych w warstwach splotu. Do szkolenia wykorzystano framework Caffe. Można zaobserwować, że wykorzystanie nowszej architektury GPU dało rezultaty około dwukrotnego przyspieszenia czasu szkolenia w przypadku AlexNet i ResNet oraz około 30% w przypadku GoogLeNet. AlexNet zawiera najmniej warstw konwolucyjnych i optymalizowaną implementację klasyfikatora. Szkolenie tej architektury zajęło mniej niż jedną godzinę na GPU Volta. Dla porównania, sieci z bardziej skomplikowaną topologią tj. ResNet-18 i GoogLeNet, charakteryzują się odpowiednio czasami szkolenia w okolicach 20 i 48 godzin.

Biorąc pod uwagę zestawienie wyników dokładności klasyfikacji, AlexNet i GoogleNet charakteryzują się tendencją do większej generalizacji. Z tych dwóch modeli, czas szkolenia jest o blisko 47 godzin krótszy dla AlexNet, dlatego to właśnie ekstraktor cech tej architektury został wybrany do realizacji kolejnych eksperymentów.

Ostatecznie, należy również podkreślić, że istnieją sposoby na dalsze poprawienie wyników dokładności wnioskowania. Dla przykładu, uzupełniające badanie dotyczące możliwości fuzji sieci AlexNet i ResNet-18 zostało opublikowane przez autora tej pracy

w czasopiśmie Acta of Bioengineering and Biomechanics [95]. Wyniki mogą być również poprawione poprzez analizę lokalizacji przekrojów w przestrzeni trójwymiarowej. W szczególności poprzez uwzględnienie, że przekroje oddalone od miejsca zerwania bardziej przypominają tkanki zdrowe. Można zatem np. zastosować metody logiki rozmytej lub miękkiej klasyfikacji (zob. []) w celu lepszego określenia przynależności próbek do klas. Niemniej jednak, z uwagi na wysokie wyniki dokładności zaprezentowanego podejścia, na tym etapie autor nie decyduje się na dalsze usprawnienia kosztem komplikacji algorytmu. Ewentualne wykryte anomalie i wyniki odosobnione eliminowane są w tej pracy poprzez zastosowanie średniej trymowanej w zaproponowanej metryce H (zob. rów. 5.1).

5.2.2 Obliczanie i redukcja wymiarowości wektora cech

Celem tego eksperymentu jest znalezienie zredukowanego wektora cech w pochodzącego z wyjścia sieci neuronowej. W założeniu ma on zawierać możliwie dużo informacji z oryginalnego zbioru i minimalny wymiar. W tym celu, z oryginalnej architektury AlexNet została wydzielona część ekstraktora cech oraz pierwsza warstwa gęsta (zob. Rys. 4.8), co umożliwiło wstępne pogrupowanie cech obrazowych, a jednocześnie brak ich dywersyfikacji w stopniu jaki robi to pełna sieć. Parametry ekstraktora cech przeniesiono z modelu wybranego w eksperymencie binarnej klasyfikacji.

Wykorzystując zbiór danych PCA, została przeprowadzona analiza czynników głównych (zob. 4.2.2), gdzie pojedynczą obserwację stanowił wektor 4096 aktywacji na wyjściu zmodyfikowanej architektury AlexNet. Analiza umożliwiła redukcję wymiarowości i rozróżnienie między obserwacjami w poszczególnych krokach czasowych. Porównanie poziomu wariancji zachowywanej przez różną liczbę czynników głównych zebrano w Tab. 5.2. Wyjścia po transformacji PCA nazywane są dalej w pracy *cechami DL*.

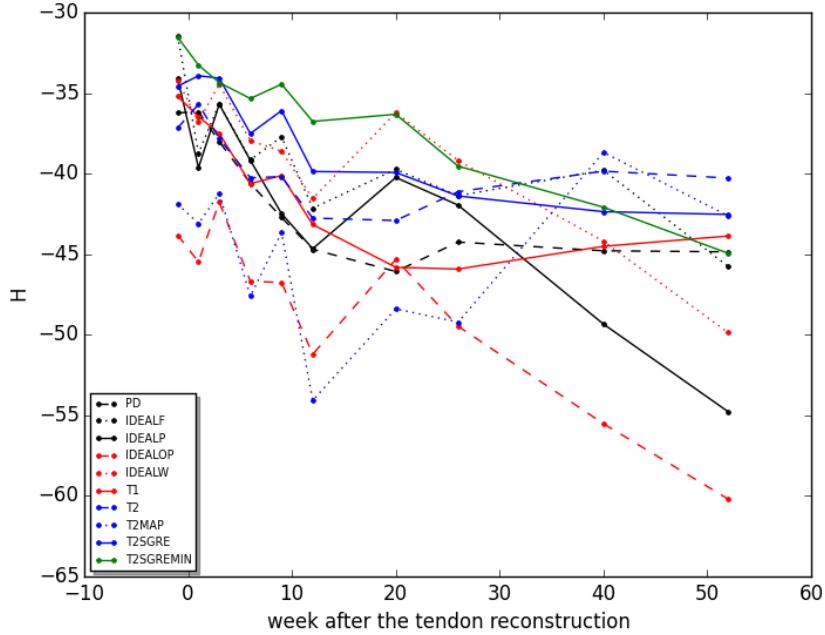
Tabela 5.2: Rezultaty PCA dla 1, 10 i 200 pierwszych czynników głównych.

	PC1	PC1-PC10	PC1-PC200
Zachowany poziom wariancji	50.2%	90.8%	98.8%

Można zaobserwować, że 4096 wartości może być zredukowanych do jednej cechy DL zachowując przy tym ponad 50% poziomu wariancji. Odpowiednio 10 cech DL zapewnia poziom powyżej 90%, a 200 poziom bliski 99%.

W kolejnym kroku zwizualizowano uśrednione wartości PC1 dla wyników wszyst-

kich pacjentów ze zbioru PCA w kolejnych tygodniach rehabilitacji (zob. Rys. 5.4).



Rysunek 5.4: Porównanie krzywych dla 10 modalności RM reprezentujących średnią H_{PC} z 10 pacjentów.

Oddziennie przedstawiono wykresy uzyskanych krzywych dla wszystkich użytych w tej pracy sekwencji i modalności RM. W celu uzyskania pojedynczej wartości opisującej badanie RM wykorzystano ponownie średnią trymowaną:

$$H_{PC} = TM(PC1(x_1), PC1(x_2), \dots, PC1(x_n)) \quad (5.2)$$

Wizualizacja umożliwia lepsze zrozumienie informacji zachowywanej przez użytą transformację PCA. Oprócz sekwencji T2 mapping, wszystkie pozostałe pokazują malejący trend rozumiany jako ujemna różnica pomiędzy wartością końcową i początkową okresu rehabilitacji. Odmienny trend T2 mapping może być spowodowany faktem, że dla tej sekwencji dane były zbierane we wszystkich 8 różnych poziomach sygnału T2. W związku z czym, w obrazach pojawiają się próbki silnie zaszumione i zniekształcone. Dopiero głębsza analiza związana np. z wyliczeniami na podstawie aproksymacji wartości tych poziomów może skutkować efektywnym wykorzystaniem T2 mapping do oceny zadanego w tej pracy problemu (zob. [103]).

W pozostałych przypadkach otrzymane trendy odpowiadają naturze procesu goje-

nia i poszczególnym jego etapom. Na początku procesu zmiany są bardzo wyraźne w badaniach obrazowych, co zmniejsza się na kolejnych etapach. Dynamika i fluktuacje zależą również od indywidualnych cech pacjenta oraz innych czynników jak przestrzeganie zaleceń rehabilitacji, diety i aktywności pacjenta.

Podsumowując, użyta transformacja pozwala na uzyskanie zredukowanego wektora cech DL zawierającego od 1 do 200 wartości i zachowującego od 50.2–98.8% informacji. Średnie trendy tak otrzymywanych wartości wskazują również na możliwość wnioskowania, co do procesu gojenia się ściegna.

5.2.3 Wybór sekwencji rezonansu magnetycznego

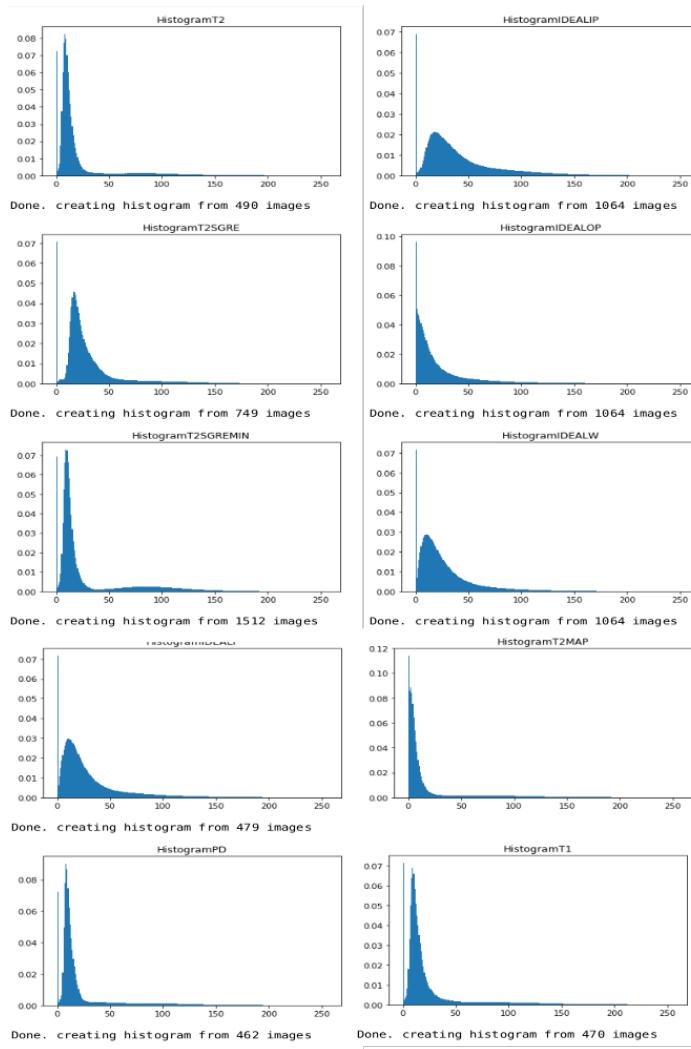
Z uwagi na opisany w Sekcji 3.2 tor akwizycji danych RM, zwiększanie liczby sekwencji odbywa się kosztem wydłużenia czasu badania pacjenta. W skrajnym przypadku, takim jak pozyskanie wszystkich 10 omawianych sekwencji, czas ten wynosi około 1 godziny, co stanowi praktyczne ograniczenie opracowywanej metody. Dlatego celem tego eksperymentu jest wybór protokołu zawierającego minimalną liczbę sekwencji umożliwiających przy użyciu opracowywanej metody skuteczną ocenę procesu gojenia. Prace podzielono na dwa etapy tj. analizę wizualną i ilościową.

Analiza wizualna:

W ramach wstępnej analizy, wartości danych ze wszystkich sekwencji i modalności zebranych od pacjentów zostały zakumulowane w postaci histogramów i przedstawione na Rys. 5.5. Dane zostały przeskalowane do zakresu $<0; 255>$ i zmodyfikowane do porównań tak, aby wartość całki Riemanna była równa 1.

Widoczne we wszystkich przypadkach maksima globalne w zerze są naturalnie wynikiem tła, które stanowi powietrze. Analizując pozostałe wartości można wyróżnić następujące 3 grupy:

1. Bez maksimum lokalnego – sekwencja T2 mapping i modalność Out Phase Ideal. W pierwszym przypadku jest to konsekwencja wspomnianej już w poprzednim eksperymencie występowania danych ze wszystkich 8 próbek sekwencji, czyli również tych mierzonych przy prawie zerowym sygnale T2, silnie zaszumionych z licznymi wartościami bliskimi zeru. W drugim przypadku na granicach wody i



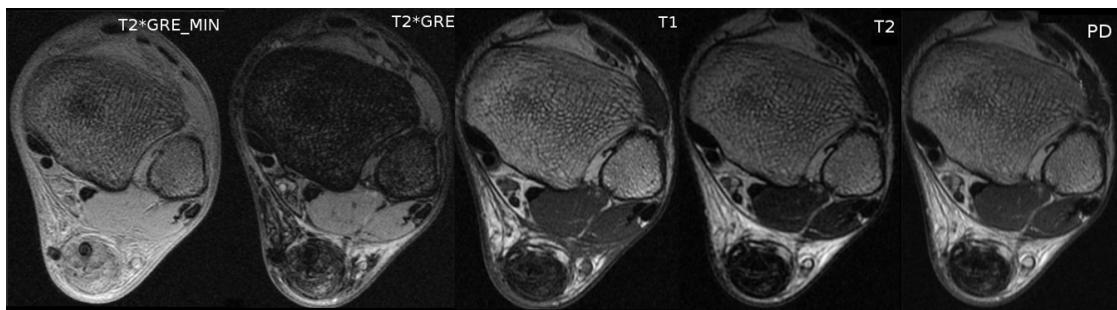
Rysunek 5.5: Znormalizowane histogramy dla 10 sekwencji RM.

tłuszcza występuje często silne rozmycie (tzw. *artefakt czarnego atramentu*), co również zwiększa liczbę wartości bliskich zeru.

2. Z jednym maksimum lokalnym – pozostałe modalności 3D FSPGR oraz sekwencje: PD, T1 zależne, T2 zależne, T2 * GRE. Przypadek najczęściej występujący. Wartości sygnału zależą od tkanek dominujących w obrazowanym elemencie.
3. Z dwoma maksimami lokalnymi – sekwencja T2 * GRE TE_MIN. Jako jedyna posiada drugie maksimum lokalne, co może wskazywać na zwiększoną czułość na wzorce zawarte w danych.

Na podstawie powyższej, zgrubnej analizy można wyłączyć z dalszych badań pierwszą

grupę sekwencji, z uwagi na możliwość występowania wysokiego szumu w sygnale oraz artefaktów. Zdecydowanie wyróżniającą się sekwencją jest T2 * GRE TE_MIN. Żeby lepiej zrozumieć istotność występowania drugiego maksimum na Rys. 5.6 zestawiono przykład rekonstrukcji danych z wykorzystaniem sekwencji z obu analizowanych dalej grup (wyłączono sekwencję 3DFSPGR, której modalności podzieliły się pomiędzy 1 i 2 grupę oraz, która we wstępnych badaniach charakteryzowała się dużym zaszumieniem).

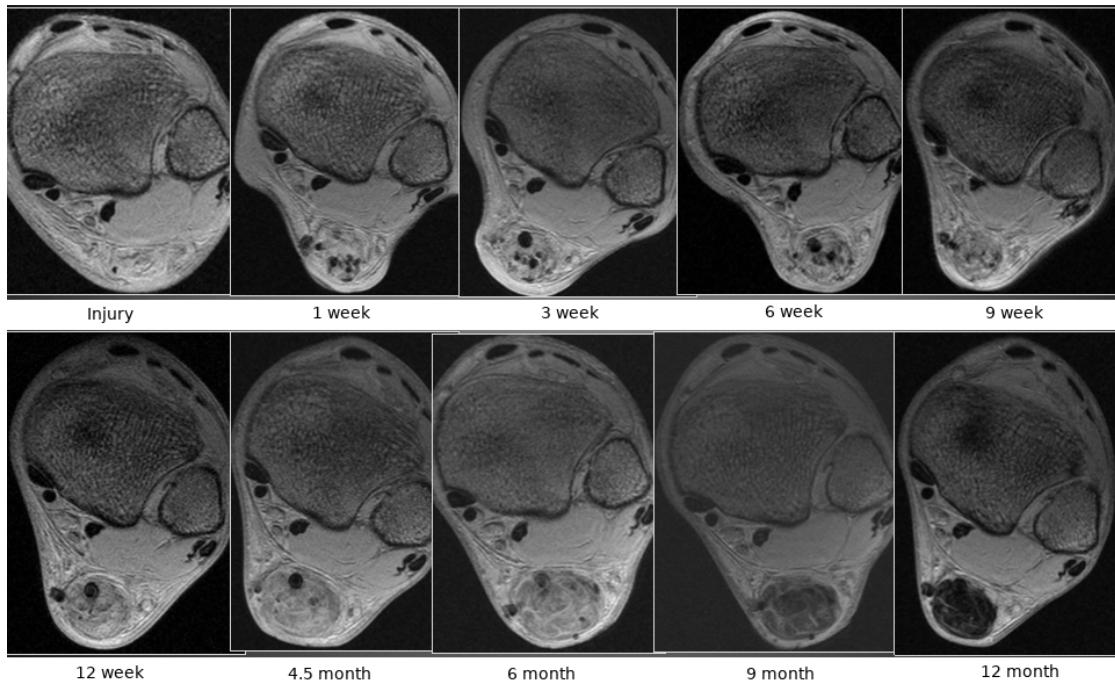


Rysunek 5.6: Porównanie sekwencji RM ilustrujących proces gojenia się ścięgna Achillesa w 12 tygodniu po rekonstrukcji.

Główna różnica w rekonstruowanych przy pomocy powyższych danych obrazów pochodzi z poziomów odcieni szarości w obszarze ścięgna. Obrazy w dwunastym tygodniu procesu gojenia się, rekonstruowane przy pomocy sekwencji z grupy (2), charakteryzują się wartościami w obszarze ścięgna bliskimi zeru. Jedynie na obrazach rekonstruowanych z wykorzystaniem danych z T2 * GRE TE_MIN obszar ścięgna jest jaśniejszy.

Na Rys. 5.7 przedstawiono próbki T2 * GRE TE_MIN zrekonstruowane w kolejnych krokach czasowych.

Analiza wizualna wskazuje, że różnice w odcieniach szarości mogą wspomóc radiologów w skutecznej interpretacji stopnia wygojenia się tkanek ścięgnistych. Wartości w obszarze ścięgna są jasne na początku i ciemnieją z czasem. Fakt ten może wynikać z czasu relaksacji T2, który dla sekwencji T2 * GRE TE_MIN jest wyjątkowo krótki. Stąd sekwencja jest bardzo czuła na ilość krwi dostarczanej do tkanek. Zmiany w tym zakresie, na przestrzeni całego procesu gojenia się ścięgna uwidaczniają się w postaci silnie rozróżnialnych zmian w odcieniach szarości w obszarze ścięgna, co z kolei jest również wartościową informacją z uwagi na zastosowane w tej pracy metody sztucznej inteligencji i przetwarzania obrazów. Powyższe obserwacje i wnioski zostały dalej wykorzystane w analizie ilościowej.



Rysunek 5.7: Proces gojenia się ścięgna Achillesa widoczny w obrazach zrekonstruowanych na podstawie danych z sekwencji T2 * GRE TE_MIN

Analiza ilościowa:

W ramach tego badania, ilościowo porównano wyniki automatycznej oceny procesu gojenia się ścięgna Achillesa przy użyciu danych (1) tylko z sekwencji T2 * GRE TE_MIN; (2) danych z sekwencji T2 * GRE TE_MIN oraz PD i T2 * GRE. Wybierając sekwencję T2 * GRE TE_MIN kierowano się wynikami analizy wizualnej, natomiast dobór pozostałych dwóch sekwencji, poza analizą wizualną i wiedzą dziedzinową, był umotywowany oceną stopnia korelacji zaprezentowaną w Tab. 5.3, gdzie zestawiono sekwencje z grupy (2) bez 3DFSPGR.

Tabela 5.3: Korelacja sekwencji PD, T1, T2 and T2* GRE MRI. Wyniki oznaczone pogrubieniem są istotne statystycznie z $p < 0.01$.

	PD	T1	T2	T2*GRE
PD	1.00	0.96	0.90	0.89
T1	0.96	1.00	0.85	0.92
T2	0.90	0.85	1.00	0.71
T2*GRE	0.89	0.92	0.71	1.00

Do obliczeń korelacji wybrano wektor cech DL zredukowany do pojedynczej wartości

niosącej ponad 50% informacji, zagregowany dla badania z wykorzystaniem metryki H_{PC} (zob. Row. 5.2).

Uwzględniając wyniki istotne statystycznie, wśród najbardziej skorelowanych sekwencji znalazły się T1 i PD. PD najmniej koreluje z T2*GRE, natomiast T1 z sekwencją T2.

W kolejnym kroku wykorzystano te informacje sprawdzając, w odniesieniu do wybranej w analizie wizualnej sekwencji T2 * GRE TE_MIN, czy wnioskowanie, co do procesu gojenia, na podstawie cech DL ulega polepszeniu z wykorzystaniem większej liczby różnorodnych sekwencji.

W celach porównań, wykorzystano ponownie zredukowaną przestrzeń cech DL (tym razem od 1–10 czynników głównych) i wykonano szereg obliczeń stopnia gojenia się ścięgna stosując do obliczenia pojedynczej wartości per badanie poniższą metrykę:

$$H_{reg} = \alpha + \sum_{i=1}^3 \beta_i X_i + \sum_{i=1}^3 \gamma_i X_i^2 + \sum_{\substack{i,j=1 \\ i < j}}^3 \lambda_{i,j} X_i X_j \quad (5.3)$$

gdzie $X_i = TM(PC_n(x_1), PC_n(x_2), \dots, PC_n(x_n))_i$ to kolejne predyktory, TM to średnia trymowana z marginesami 2.5%, $PC_n(x_k)$ to n -ty czynnik główny otrzymany przy wnioskowaniu sieci dla przekroju osiowego x_k , gdzie k jest indeksem przekroju danej sekwencji w trójwymiarowym badaniu RM.

W szczególności, wyliczono różne modele regresji poczynając od liniowej, gdzie ($\gamma_i = 0, \lambda_i = 0$), przez nieliniową stopnia drugiego ($\lambda_i = 0$) oraz z kombinacją czynników. W przypadku liniowym wykorzystano od 1 do 10 czynników głównych. Natomiast w przypadku nieliniowym wykorzystano tylko dwa pierwsze czynniki, ograniczając tym samym efekt nadmiernego dopasowania się modelu. Powyższe modele funkcjonujące w oparciu o trzy sekwencje (tj. T2 * GRE TE_MIN, PD i T2 * GRE) zestawiono z modelem liniowej regresji wytrenowanym jedynie na danych z T2 * GRE TE_MIN. Do obliczenia współczynników regresji zastosowano zbiór pacjentów treningowych. Natomiast w celach porównań zestawiono wyniki dla pacjentów testowych. W szczególności, wyliczono metryki średniego błędu absolutnego (MAE od ang. *Mean Absolute Error*), maksymalnego błędu absolutnego (MAX-AE od ang. *Max. Absolute Error*) i średniej korelacji obliczonej z wykorzystaniem transformacji Z-fishera (Corr od ang. *Correlation*). Rezultaty znajdują się w Tab. 5.4.

Model prostej liniowej regresji wykorzystujący jedynie informacje z sekwencji T2 *

Tabela 5.4: Porównanie wnioskowania dla różnych modeli regresji wytrenowanych w oparciu o dane z 3 sekwencji RM, z modelem liniowej regresji wytrenowanym w oparciu o jedną sekwencję RM.

Model		SCT	TT	STE	TE	TU	TisE
model_PC1	MAE	1.26 ± 0.44	0.71 ± 0.12	0.7 ± 0.16	0.97 ± 0.26	0.92 ± 0.29	0.99 ± 0.25
	MAX-AE	3.53	2.49	1.91	2.34	2.2	2.47
	Corr	0.58	0.47	-0.07	0.60	0.56	0.58
model_PC1_2	MAE	1.28 ± 0.47	0.74 ± 0.13	0.71 ± 0.19	0.98 ± 0.26	0.94 ± 0.3	1.02 ± 0.28
	MAX-AE	3.64	2.56	1.94	2.98	2.35	2.9
	Corr	0.53	0.30	0.02	0.51	0.41	0.60
model_PC1_3	MAE	1.43 ± 0.58	0.71 ± 0.13	0.73 ± 0.23	1.02 ± 0.31	1.05 ± 0.36	1.07 ± 0.37
	MAX-AE	3.99	2.49	1.98	3.2	2.96	3.02
	Corr	0.48	0.36	0.03	0.50	0.29	0.62
model_PC1_4	MAE	1.39 ± 0.53	0.72 ± 0.09	0.8 ± 0.22	1.01 ± 0.3	1.02 ± 0.33	1.06 ± 0.37
	MAX-AE	3.56	2.4	2.17	3.2	2.61	3.13
	Corr	0.54	0.37	0.17	0.52	0.38	0.68
model_PC1_5	MAE	1.35 ± 0.5	0.79 ± 0.06	0.8 ± 0.25	0.99 ± 0.26	1.01 ± 0.26	1.1 ± 0.29
	MAX-AE	3.45	2.69	2.39	3.21	2.55	2.88
	Corr	0.49	0.31	0.22	0.52	0.30	0.60
model_PC1_6	MAE	1.32 ± 0.49	0.78 ± 0.05	0.8 ± 0.25	0.97 ± 0.29	0.98 ± 0.24	1.11 ± 0.25
	MAX-AE	3.3	2.64	2.21	3.12	2.6	2.8
	Corr	0.53	0.33	0.21	0.56	0.38	0.56
model_PC1_7	MAE	1.33 ± 0.5	0.78 ± 0.07	0.82 ± 0.26	1.03 ± 0.31	0.99 ± 0.26	1.14 ± 0.29
	MAX-AE	3.34	2.75	2.41	3.17	2.63	2.69
	Corr	0.53	0.33	0.21	0.54	0.30	0.57
model_PC1_8	MAE	1.35 ± 0.49	0.77 ± 0.08	0.81 ± 0.26	1.02 ± 0.31	0.99 ± 0.28	1.14 ± 0.27
	MAX-AE	3.41	2.58	2.47	3.13	2.65	2.75
	Corr	0.54	0.37	0.19	0.53	0.25	0.55
model_PC1_9	MAE	1.38 ± 0.48	0.74 ± 0.09	0.76 ± 0.27	1.03 ± 0.28	1.02 ± 0.3	1.12 ± 0.27
	MAX-AE	3.49	2.57	2.45	2.9	2.7	2.67
	Corr	0.52	0.40	0.25	0.55	0.30	0.58
model_PC1_10	MAE	1.37 ± 0.49	0.74 ± 0.11	0.76 ± 0.27	1.04 ± 0.27	1.03 ± 0.29	1.11 ± 0.28
	MAX-AE	3.46	2.42	2.52	2.92	2.67	2.68
	Corr	0.50	0.38	0.24	0.54	0.29	0.59
model_poly	MAE	1.27 ± 0.49	0.74 ± 0.1	0.71 ± 0.2	0.93 ± 0.28	0.95 ± 0.31	1.03 ± 0.25
	MAX-AE	3.62	2.6	1.96	2.83	2.29	2.87
	Corr	0.53	0.34	0.11	0.61	0.37	0.58
model_cross_poly	MAE	1.3 ± 0.47	0.75 ± 0.1	0.78 ± 0.24	0.93 ± 0.28	1.01 ± 0.42	1.06 ± 0.32
	MAX-AE	3.64	2.7	2.01	2.83	3.36	2.86
	Corr	0.53	0.45	0.12	0.58	0.15	0.57
baseline	MAE	1.24 ± 0.16	0.82 ± 0.09	0.75 ± 0.08	1.06 ± 0.10	0.90 ± 0.09	0.96 ± 0.10
	MAX-AE	3.54	2.46	1.82	2.70	2.13	2.18
	Corr	0.61	0.64	-0.08	0.55	0.55	0.65

GRE TE_MIN był najskuteczniejszy w największej liczbie mierzonych parametrów tj. 8 z 18. 3 razy osiągnął najlepszy wynik w MAE i MAX-AE oraz dwukrotnie w Corr. Zatem dodawanie sekwencji jako danych wejściowych i komplikacje modelu skutkują nadmiernym dopasowaniem oraz pogorszeniem jakości automatycznej oceny. Należy również podkreślić, że akwizycja samej sekwencji T2 * GRE TE_MIN dla omawianego problemu zajmuje około 5 min., podczas gdy wszystkich trzech sekwencji trwa około 12 minut na pacjenta. Podsumowując, na podstawie analizy wizualnej i jakościowej, jako dane wejściowe do kolejnych eksperymentów została wybrana sekwencja T2 * GRE

TE_MIN.

5.2.4 Optymalna selekcja i konkatenacja cech

W poprzednim eksperymencie uzyskano najlepsze wyniki z wykorzystaniem modeli liniowej regresji. Można zatem postawić hipotezę, że w ocenie radiologa dominuje zależność liniowa. Dodatkowo, cechy obrazowe opisujące jeden problem diagnostyczny są ze sobą zazwyczaj silnie skorelowane. Stąd motywacja, aby w ramach tego eksperimentu do selekcji cech zastosować metodę LASSO (zob. pkt 4.2.2), bazującą właśnie na liniowych zależnościach i znajdującą swe zastosowanie przy silnie skorelowanych predyktorach.

W tym celu, treningowy zbiór pacjentów został podzielony na 4 podzbiory (11 pacjentów i około 4300 przekrojów osiowych) i wykorzystany następnie do kroswalidacji. Wynikiem tych działań był dobór możliwie najlepszego parametru α dla metody LASSO. Jako dane wejściowe wykorzystano zbiór ograniczony do sekwencji T2 * GRE TE_MIN, z których dla każdego przekroju wyliczono odpowiednio 200 cech DL i 46 klasycznych cech obrazowych (zob. 5.1).

Kryterium selekcji cech była najlepsza korelacja z wzorcem odniesienia, którą uzyskano dla $\alpha = 0.1$. W rezultacie dla 6 parametrów wzorca odniesienia uzyskano następujące podzbiory predyktorów:

- SCT: 6 cech DL (PC1, PC2, PC3, PC5, PC7, PC9), 10 klasycznych cech obrazowych, w tym 3 cechy Haralicka (suma wariancji dla $d = 1$, suma średnich dla $d = 5$ i suma średnich dla $d = 10$);
- TT: 5 cech DL (PC1, PC2, PC3, PC4, PC5) i 7 klasycznych cech obrazowych bez udziału cech Haralicka;
- STE: 5 cech DL (PC1, PC4, PC7, PC8, PC9) i 8 klasycznych cech obrazowych bez udziału cech Haralicka;
- TE: 4 cechy DL (PC1, PC2, PC3, PC9) i 8 klasycznych cech obrazowych, w tym jedna cecha Haralicka tj. suma średnich dla $d = 10$;
- TU: 4 cechy DL (PC1, PC3, PC4, PC9) i 9 klasycznych cech obrazowych, w tym jedna cecha Haralicka tj. suma wariancji dla $d = 1$;

- TisE: 6 cech DL (PC1, PC2, PC3, PC5, PC7, PC9) i 7 klasycznych cech obrazowych bez udziału cech Haralicka.

Maksymalna liczba cech w zredukowanej przestrzeni wyniosła 16 dla parametru SCT. Ponadto cechy DL i klasyczne cechy obrazowe są obecne w każdym z ocenianych kryteriów wzorca odniesienia, co wskazuje na możliwość uzyskania polepszonych rezultatów oceny z wykorzystaniem fuzji. Należy również zwrócić uwagę, że największa liczba cech Haralicka (tj. 3) występuje w parametrze SCT związanym z oceną struktury włókien widoczną w obrazach jako różne tekstury.

Wyselekcjonowane podzbiory wykorzystano w kolejnym kroku eksperymentu mającym na celu porównanie oceny procesu gojenia z wykorzystaniem różnych metod. Wykorzystano zarówno liniową regresję jak i tym razem algorytmy mogące modelować nieliniowe zależności. Pomimo wstępnego założenia o dominacji liniowych związków np. wynik modelu *model_poly* dla Corr w parametrze TE (zob. 5.4) pozwala zakładając, że ocena w niektórych detalach może mieć również charakter nieliniowy.

Dlatego porównano różne metody (liniowe i nieliniowe) finalnej meta-regresji cech tj.: metodę wektorów nośnych (SVR – od ang. *Support Vector Regression*) opisaną w [X], wielowarstwowy perceptron liniowy z 4 neuronami w warstwie ukrytej (MPR – od ang. *Multilayer Perceptron Regression*) oraz las losowy składający się ze 100 drzew decyzyjnych (RF – od ang. Random Forest). Wyniki zestawiono w Tab. 5.5 z rezultatami liniowej regresji (LR – od ang. Linear regression) i regresji nieliniowej drugiego stopnia (Poly – od ang. Polynomial). Trening metod i dobór hiperparametrów wykonano z wykorzystaniem kroswalidacji z podziałem na 4 segmenty i zbioru treningowych pacjentów. Wyniki oceny zestawiono w Tab. 5.5.

Ponownie wykorzystano metryki średniego błędu absolutnego (MAE), maksymalnego błędu absolutnego (MAX-AE) i średniej korelacji obliczonej z wykorzystaniem transformacji Z-fishera (Corr). Najlepszy rezultat dla każdego z parametrów uzyskano dla lasu losowego, jednak w przypadku tego algorytmu (pomimo wielu prób z różną liczbą drzew decyzyjnych) występuje efekt nadmiernego dopasowania, co można zaobserwować analizując wyniki ze zbioru testowego zestawione w Tab. 5.6.

W zestawieniu z oceną bazującą tylko na cechach DL (baseline limited). W tym porównaniu nawet wyniki uzyskane z modelu liniowej regresji bazującej na fuzji cech pozwalają na znaczącą poprawę rezultatów w każdym z parametrów w przypadku korelacji, w parametrach, SCT, TT, STE i TU dla MAX-AE oraz SCT, TT, i TE dla MAE. Wśród wszystkich oprócz RF algorytmów, należy zwrócić uwagę, że metodą

Tabela 5.5: Wyniki oceny procesu gojenia z wykorzystaniem fuzji cech i zbioru treningowego.

Model		SCT	TT	STE	TE	TU	TisE
Poly	MAE	1.00 ± 0.03	0.62 ± 0.02	0.73 ± 0.02	0.76 ± 0.02	0.89 ± 0.03	0.70 ± 0.02
	MAX-AE	3.53	2.35	3.62	2.49	2.90	2.64
	Corr	0.87	0.82	0.46	0.80	0.65	0.87
SVR	MAE	0.88 ± 0.01	0.59 ± 0.01	0.67 ± 0.01	0.69 ± 0.01	0.83 ± 0.01	0.63 ± 0.01
	MAX-AE	3.73	2.32	3.83	2.50	2.95	2.75
	Corr	0.89	0.85	0.59	0.83	0.72	0.88
LR	MAE	1.00 ± 0.04	0.62 ± 0.02	0.74 ± 0.02	0.74 ± 0.02	0.90 ± 0.03	0.69 ± 0.02
	MAX-AE	3.52	2.39	3.77	2.52	2.88	2.65
	Corr	0.87	0.83	0.46	0.80	0.65	0.87
MPR	MAE	1.00 ± 0.04	0.63 ± 0.02	0.74 ± 0.02	0.71 ± 0.02	0.88 ± 0.03	0.96 ± 0.02
	MAX-AE	3.47	2.51	3.57	2.52	2.86	2.65
	Corr	0.86	0.83	0.46	0.80	0.65	0.88
RF	MAE	0.32 ± 0.04	0.21 ± 0.02	0.25 ± 0.03	0.25 ± 0.02	0.31 ± 0.03	0.23 ± 0.02
	MAX-AE	1.24	0.82	1.33	0.89	1.03	0.92
	Corr	0.99	0.99	0.99	0.99	0.99	0.99

meta-regresji, która uzyskała sumarycznie najwięcej najlepszych wyników Corr, MAE i MAX-AE okazała się SVR, co może sugerować potwierdzenie występowania niewielkich nieliniowych zależności występujących w ocenie radiologa.

Fakt poprawy wyników z wykorzystaniem fuzji klasycznych cech obrazowych i cech DL wynika z explicite wkomponowania w proces automatycznej oceny predyktorów opisujących tkanki wewnętrz ROI. Są to cechy bardziej odporne na wspomniane wcześniej zaburzenia procesu gojenia, które można ogólnie nazwać szumem w sygnale gojenia (np. znaczące zwiększenie się obrzęku na skutek temperatury lub wysiłku fizycznego). W szczególności cechy Haralicka zostały uwzględnione przy analizie SCT, TE i TU. Statystyki z ROI w znaczący sposób polepszyły analizę TT, który to parametr uzyskał najlepsze wyniki metryk spośród wszystkich. Dodatkowo, zwiększoną liczbę cech DL zwiększa poziom zachowywanej informacji, co ma wpływ na ogólną ocenę jak i w szczególności na parametr TisE zależny od zewnętrznych obszarów. Jedynie parametr STE (choć wciąż znacząco poprawiony w stosunku do przyjętego poziomu odniesienia) uzyskał korelacje znacząco poniżej 0.5. Wynik ten będzie dyskutowany w dalszej części tej rozprawy.

Dodatkowo, dla lepszego zrozumienia różnic między obszarami objętymi cechami DL i z ROI wykonano szereg eksperymentów i analiz w celu wizualizacji obszaru zainteresowania sieci neuronowych. Przykład znajduje się na Rys. 5.8. Obszar zainteresowania wykorzystanej w eksperymentach sieci neuronowej obejmuje ROI jak i tkanki okalające. Optymalna fuzja cech z widocznego obszaru ocenianego przez sieć neurono-

Tabela 5.6: Wyniki oceny procesu gojenia z wykorzystaniem fuzji cech i zbioru testowego.

Model		SCT	TT	STE	TE	TU	TisE
Poly	MAE	1.15 ± 0.13	0.57 ± 0.07	0.75 ± 0.08	0.94 ± 0.10	0.92 ± 0.09	0.94 ± 0.10
	MAX-AE	2.67	1.78	1.81	2.50	2.12	2.39
	Corr	0.82	0.83	0.25	0.71	0.63	0.78
SVR	MAE	1.05 ± 0.12	0.56 ± 0.06	0.75 ± 0.08	0.91 ± 0.10	0.91 ± 0.09	0.94 ± 0.10
	MAX-AE	2.62	1.82	1.92	2.54	2.01	2.38
	Corr	0.85	0.85	0.31	0.72	0.65	0.80
LR	MAE	1.15 ± 0.13	0.55 ± 0.06	0.74 ± 0.08	0.90 ± 0.10	0.93 ± 0.09	0.97 ± 0.10
	MAX-AE	2.60	1.78	1.81	2.54	2.04	2.38
	Corr	0.84	0.84	0.18	0.71	0.62	0.78
MPR	MAE	1.13 ± 0.12	0.57 ± 0.06	0.74 ± 0.08	0.92 ± 0.11	0.91 ± 0.09	0.96 ± 0.10
	MAX-AE	2.63	1.77	1.78	2.54	2.04	2.40
	Corr	0.83	0.83	0.20	0.73	0.63	0.77
RF	MAE	1.12 ± 0.13	0.55 ± 0.06	0.74 ± 0.08	0.90 ± 0.11	0.93 ± 0.09	0.94 ± 0.10
	MAX-AE	2.73	1.77	1.78	2.50	2.12	2.27
	Corr	0.83	0.84	0.37	0.71	0.61	0.78
baseline limited	MAE	1.24 ± 0.16	0.82 ± 0.09	0.75 ± 0.08	1.06 ± 0.10	0.90 ± 0.09	0.96 ± 0.10
	MAX-AE	3.54	2.46	1.82	2.70	2.13	2.18
	Corr	0.61	0.64	-0.08	0.55	0.55	0.65

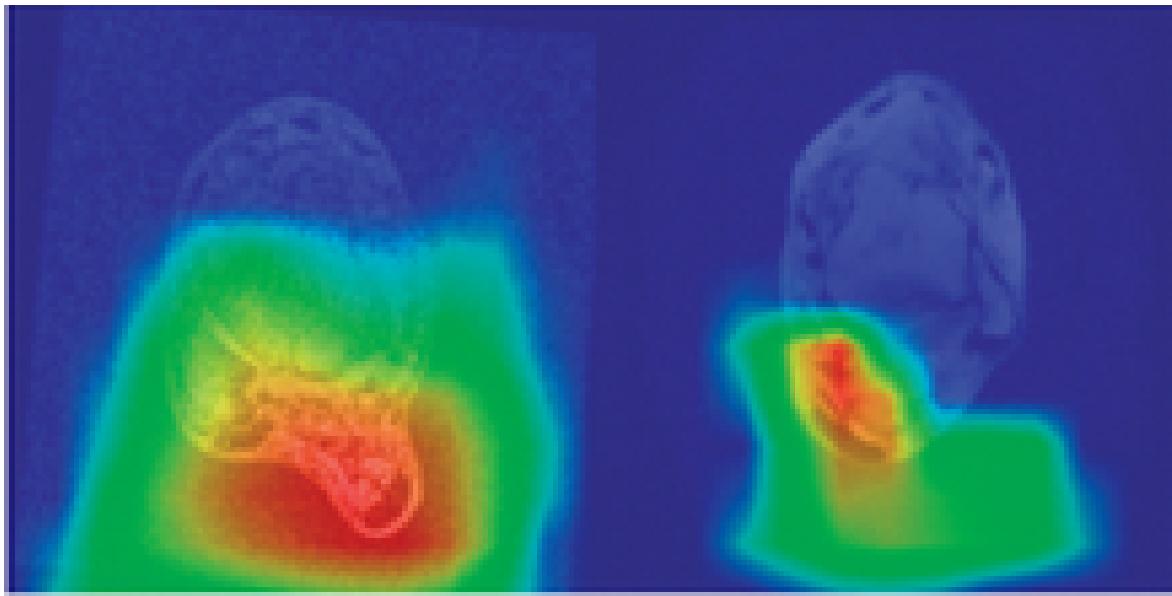
wą oraz ROI umożliwia bardziej skuteczną ocenę całego procesu poprzez zapewnienie zgrubnej oceny z użyciem sieci i szczegółowej z dołączenia predyktorów z ROI, bardziej odpornych na szum.

5.3 Ocena procesu gojenia

W ramach tego eksperymentu zostanie przedstawiona szczegółowa analiza wyników oceny procesu gojenia z wykorzystaniem proponowanej w tej pracy metody. W pierwszej kolejności dokonano zgrubnej analizy. Przeanalizowano macierze błędu dla różnych klas: (1) zakres oceny (0–7), gdzie wartością jest liczba występowania; (2) tygodnie oceny, gdzie wartością jest średni błąd absolutny (MAE). Następnie zostały przedstawione *krzywe gojenia* tj. ocena zmiany w czasie poszczególnych parametrów ankiety. Porównane zostały szczegółowo, dla wszystkich testowych pacjentów, wyniki metody automatycznej z wzorcem odniesienia tj. oceną eksperta radiologa.

5.3.1 Macierze błędu

Celem zdefiniowania różnic między oceną radiologa i propozycją automatycznej metody, zestawiono dwa rodzaje macierzy błędu. W pierwszej kolejności jako klasę



Rysunek 5.8: Analiza wizualna procesu wnioskowania sieci z zaznaczonym obszarem zainteresowania.

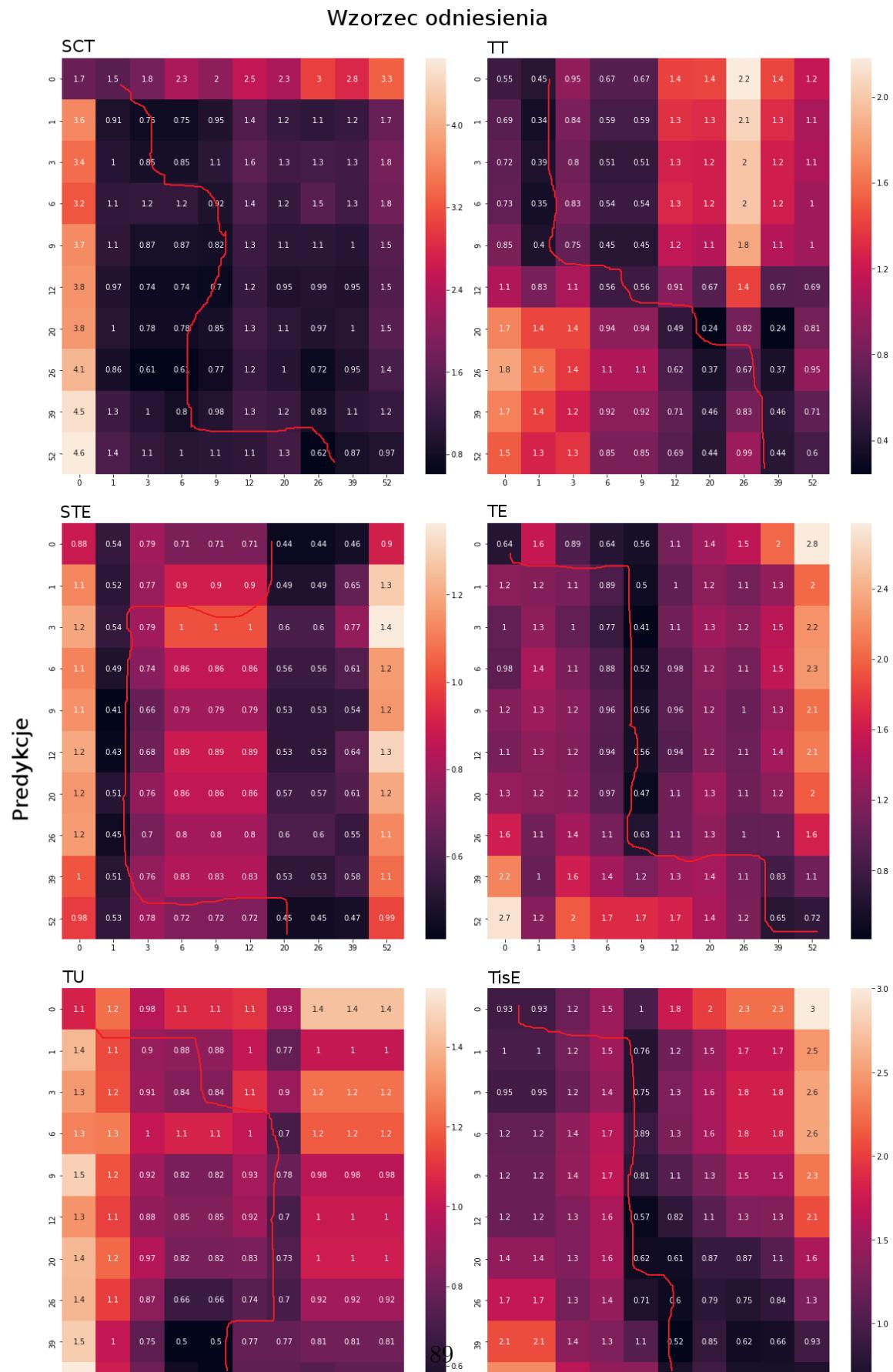
Tabela 5.7: Ocena dryfu krzywej łączącej minima lokalne wartości MAE w macierzach błędów z klasą przyjętą jako tydzień rehabilitacji.

	SCT	TT	STE	TE	TU	TisE
PP_d	18	10	42	13	30	17

wybrano tydzień gojenia się ściegna. Wyniki uzyskane z użyciem metody automatycznej porównano z oceną radiologa wykorzystując miarę średniego błędu absolutnego. Do obliczeń wykorzystano pacjentów testowych. Oddziennie dla wszystkich 6 parametrów wyznaczono macierze błędu i przedstawiono je na Rys. 5.9.

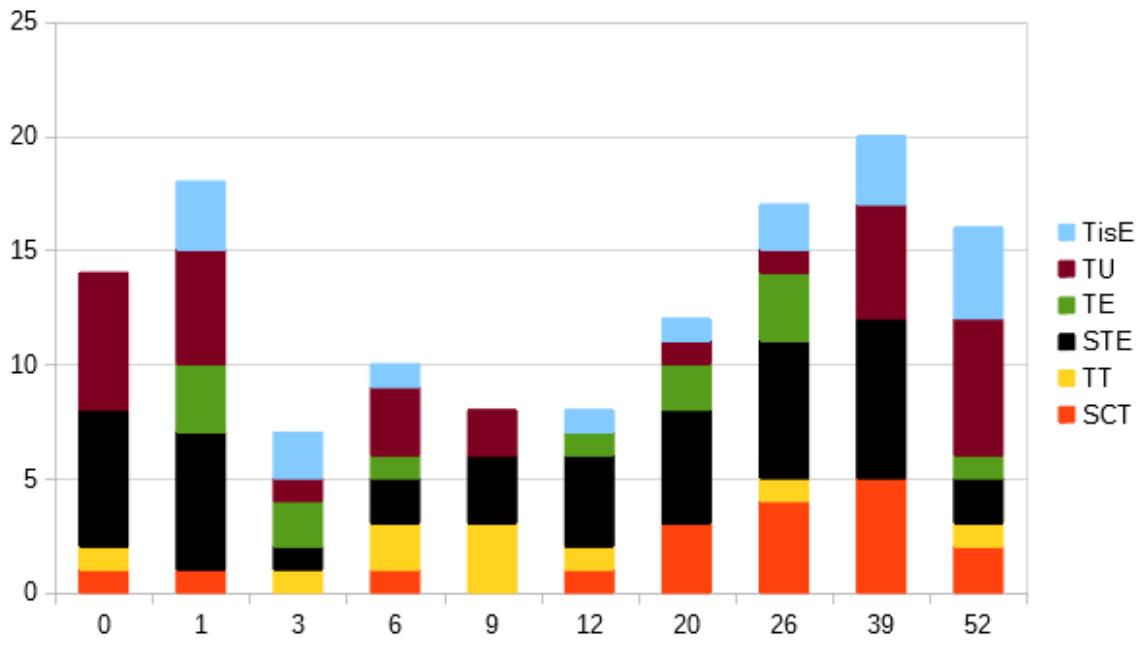
Z wykorzystaniem macierzy w stosunkowo łatwy sposób można określić w jakim stopniu wyniki ocen w poszczególnych tygodniach są do siebie zbliżone. W przypadku możliwe najlepszego odwzorowania oceny radiologa przez nową metodę, minimalne błędy znajdowałyby się na diagonali przedstawionych wykresów. Dla poszczególnych parametrów można jednak zaobserwować dryf wartości minimalnych błędów. W szczególności, pole powierzchni (PP_d) pomiędzy krzywą określającą wartości minimalne MAE a diagonalą zostały zebrane w Tabeli 5.7.

Najlepsze rezultaty (najmniejszy PP_d) uzyskano dla parametru *TT*, a najgorszy dla *STE*, co oznacza, że w tych przypadkach ocena automatyczna najlepiej i najgorzej



Rysunek 5.9: Macierz średniego błędu absolutnego dla oceny pacjentów testowych.

5.3 Ocena procesu *Pojedział 5: Nowa metoda oceny procesu gojenia ścięgna Achillesa*

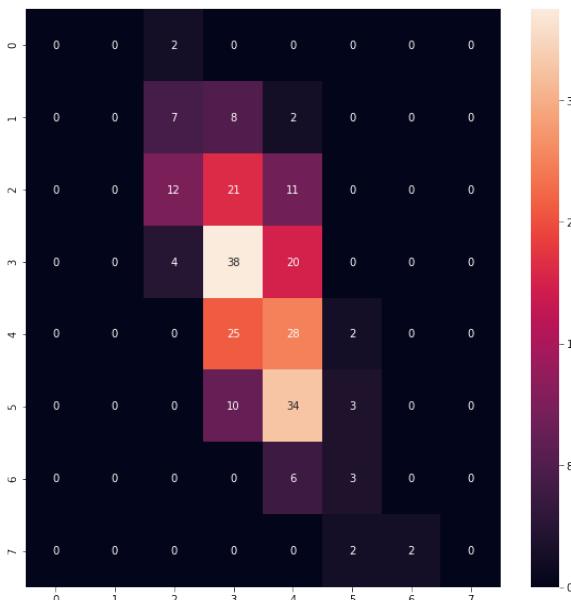


Rysunek 5.10: PP_d .

przypomina ocenę radiologa. Kolejne, najlepsze rezultaty otrzymano dla parametrów opisujących obrzęki tj. TE i $TisE$ oraz spójność ścięgna (SCT). Drugi najgorszy wynik został uzyskany dla parametru TU oceniającego jednorodność ścięgna w płaszczyźnie strzałkowej. Powyższe rezultaty są spójne z wnioskami do wyników zamieszczonych w Tab. 5.6, jednak szczegółowa analiza dryfu tygodni po tygodniu przedstawiono na Rys. 5.10 umożliwia dalsze wnioskowanie.

Najmniejsze różnice w ocenie występują w trzecim tygodniu, zatem w pierwszym pomiarze po fazie zapalnej. Największe natomiast są obecne w pierwszym oraz 39-tym tygodniu monitoringu, zatem w szczytce fazy zapalnej i w momencie zaawansowania fazy przebudowy.

Analizując poszczególne parametry od najlepszego pod względem PP_d – w przypadku parametru TT , największe różnice w ocenie widoczne są między 3–12 tyg., a zatem na przełomie fazy poliferacji i przebudowy, gdzie włókna ścięgniste zaczynają się ukierunkowywać i zmieniają grubość ścięgna. Dla TE największe odchylenia zaobserwowano dla przełomu fazy zapalnej i poliferacji – 1 tydz. oraz zaawansowanej przebudowy w 26 tyg. Podobny schemat zachowany jest dla $TisE$, z tą różnicą, że oprócz przełomu fazy zapalnej, odchylenia są widoczne w całym okresie przebudowy. Może to świadczyć o różnicy w ocenie płynów nagromadzonych w obrzękach. W przypadku



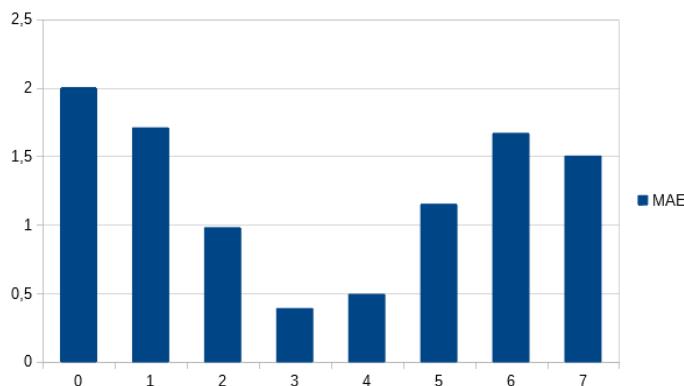
Rysunek 5.11: Macierz błędu dla noty jako klasy.

SCT, różnice w ocenie spójności włókien uwidaczniają się w fazie przebudowy, kiedy to automatyczna metoda nie wskazuje na znaczący postęp w stosunku do wyniku z 26-tego tygodnia. W przypadku dwóch parametrów uzyskujących najgorszy wynik PP_d tj. TU i STE rozbieżności są widoczne w cały okresie monitoringu, w szczególności w fazie zapalnej i przebudowy. Faza zapalna jest najbardziej dynamiczna, natomiast faza przebudowy była monitorowana w najdłuższych interwałach, stąd różnice w obrazach tkanek w obu przypadkach są największe.

Do dalszej analizy różnic między ocenami wykorzystano macierze błędu, gdzie klasą jest nota przyznawana w ocenie parametrów (zob. Rys. 5.12). Największą zgodność otrzymano dla 2, 3 i 4. Notę 5 najczęściej oceniano jako 4, podobnie jak notę 6. Notę 7, dwukrotnie oceniono jako notę 5 albo 6. W przypadku niskich wyników oceny, 1 najczęściej oceniono notą 2 (7 razy) lub 3 (8 razy). Natomiast notę 0 oceniono na wartość 2. Wartości błędów dla poszczególnych tygodni przedstawiono na Rys. 5.12.

Można zaobserwować, że najmniejsze błędy (<1) otrzymywane są w zakresie not 2–4. Dla noty 5 wynik jest nieznacznie gorszy i osiąga poziom $MAE=1,15$. W pozostałych notach skrajnych błąd wynosi $> 1,5$. Wyniki są związane z faktem, że metoda automatyczna charakteryzuje się mniejszym zakresem oceny, co wynika z zastosowania w algorytmie meta-regresji. Wartości oceny nigdy nie osiągają not skrajnych tj. 7 lub 0–1.

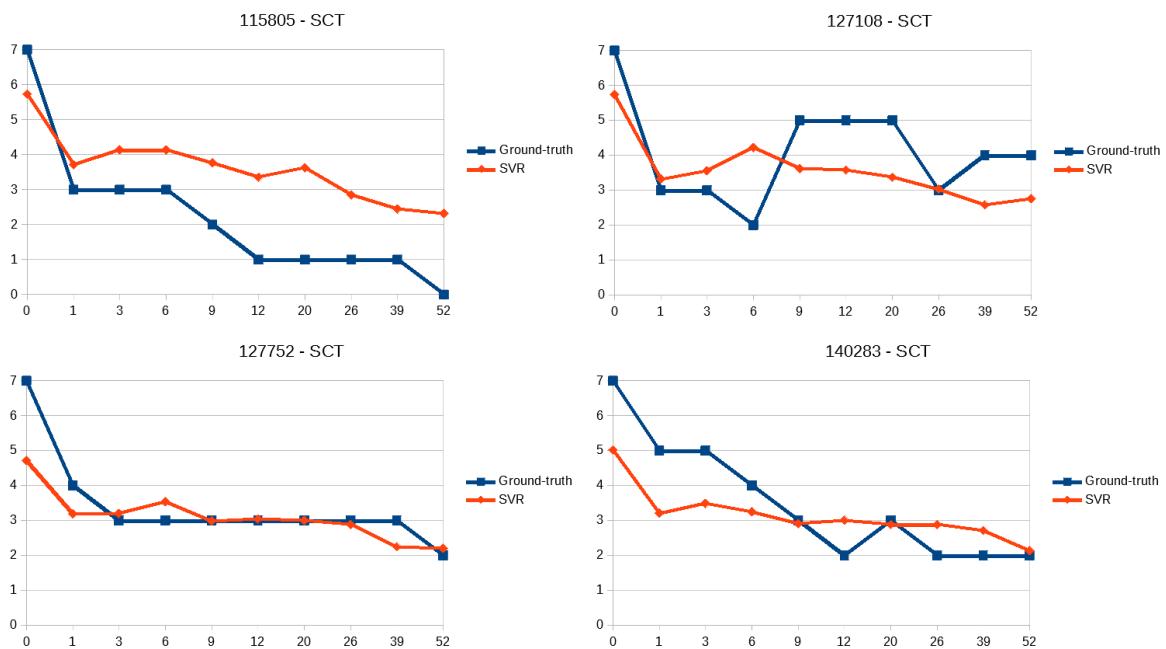
5.3 Ocena procesu *Pojedział 5: Nowa metoda oceny procesu gojenia ścięgna Achillesa*



Rysunek 5.12: MAE w obrębie klas.

5.3.2 Krzywe gojenia

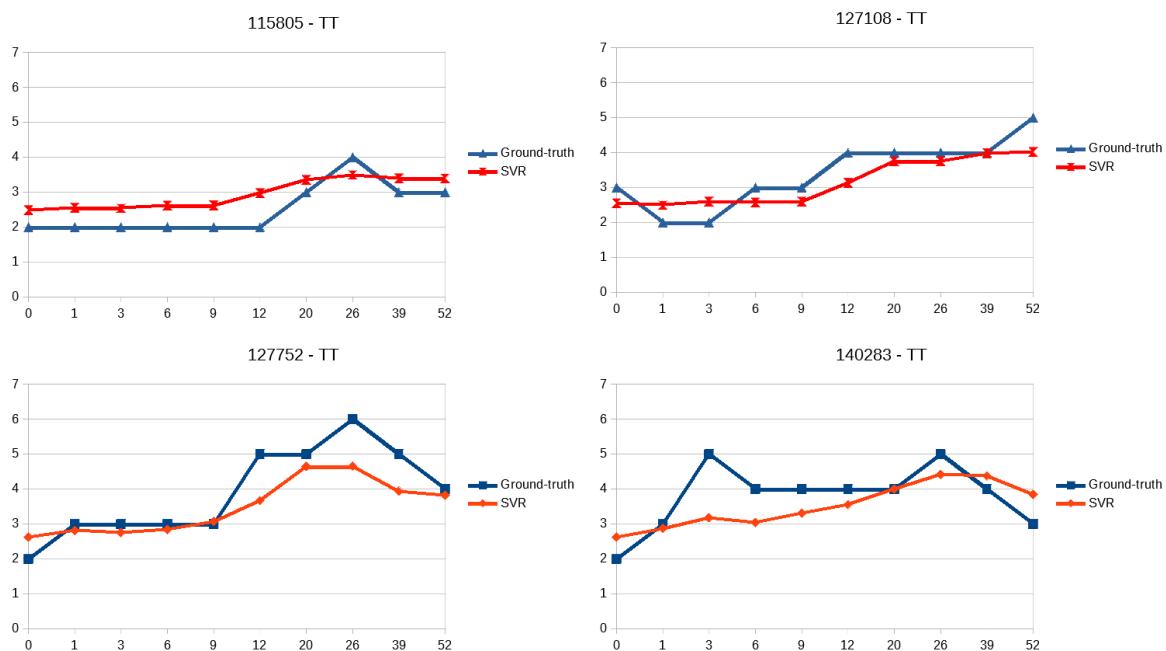
Krzywe gojenia należy w tej pracy rozumieć jako krzywą powstającą z połączenia punktów oceny automatycznej w kolejnych krokach czasowych monitorowania. W szczególności, dla wszystkich czterech pacjentów testowych, dla każdego z dziesięciu kroków czasowych zestawiono wyniki oceny automatycznej z oceną eksperta radiologa. Na kolejnych sześciu rysunkach zaprezentowano rezultaty dla wszystkich parametrów z ankiety. Jako pierwszy przedstawiono parametr SCT (zob. Rys. 5.13). Zgodnie z



Rysunek 5.13: Ocena parametru SCT.

opisem zamieszczonym w Podsekcji 5.1.2 dotyczy on stanu uszkodzeń śródścięgnistych wewnętrz obszaru ROI. Przed interwencją chirurgiczną ścięgno jest całkowicie zerwane, stąd w tygodniu 0, reprezentującym czas na kilka dni przed operacją, ocena radiologa wynosi 7, dla każdego z pacjentów. Po zszyciu w 3 z 4 przypadków proces postępuje z malejącym trendem do osiągnięcia małych, bądź braku uszkodzeń (jedyna anomalia, mogąca wynikać z omawianych wcześniej czynników zaburzających proces gojenia się, widoczna jest w 20 tyg. dla pacjenta 140283). W przypadku pacjenta 127108 przebieg gojenia się jest odmienny, przebiega ze zmiennym trendem, jak również kończy się oceną 4, interpretowaną jako dość duże uszkodzenia. Ocena automatyczna proponowana w tej pracy przybliża trendy oceny radiologa, sugerując duże lub bardzo duże uszkodzenia na początku procesu i małe uszkodzenia w 3 z 4 przypadkach na końcu. Jedynie dla pacjenta 127108 sugerowana nota końcowa oznacza uszkodzenia średniej wielkości. Poza pojedynczym maksymalnym błędem (MAX-AE) występującym w 20 tyg. w ocenie pacjenta 115805, największe obserwowalne różnice w ocenie występują przed operacją. We wszystkich czterech przypadkach ocena automatyczna stanu ekstremalnie dużych uszkodzeń związanych z całkowitym zerwaniem nie występuje, co wynika z ogólnie mniejszego zakresu punktacji metody automatycznej.

Kolejnym analizowanym parametrem jest TT. Wykresy krzywych gojenia przedstawiono na Rys. 5.14. Parametr ten zgodnie z przedstawioną interpretacją związany jest



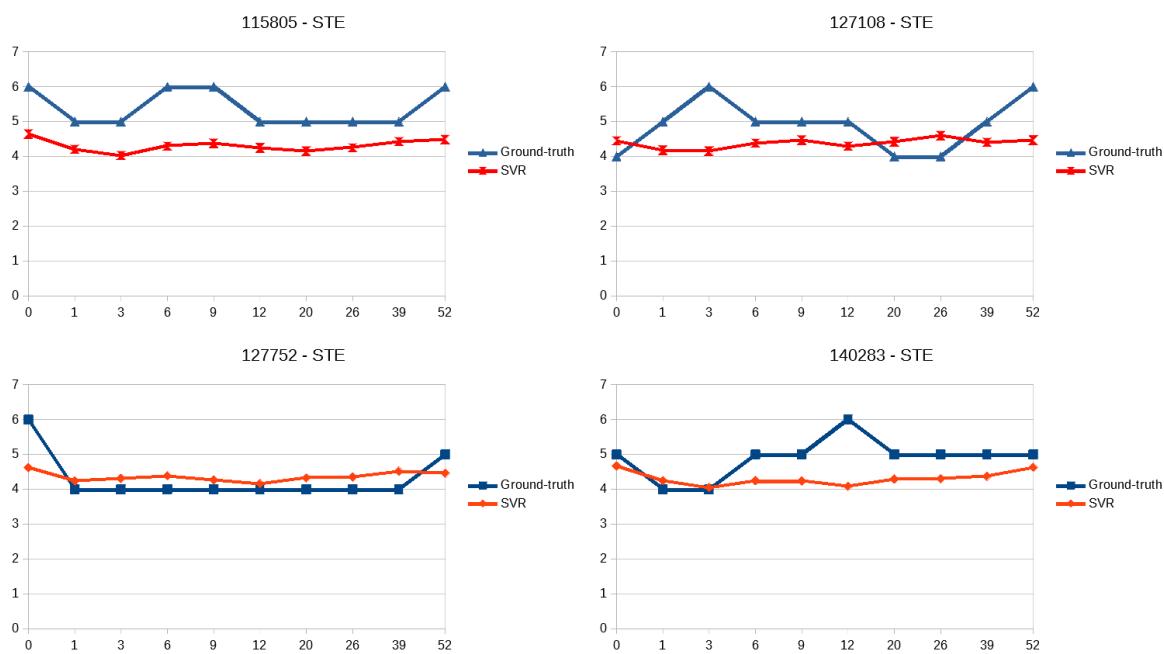
Rysunek 5.14: Ocena parametru TT.

5.3 Ocena procesu *Pojedział 5: Nowa metoda oceny procesu gojenia ścięgna Achillesa*

z oceną pogrubienia ścięgna. Na początku procesu pogrubienie jest małe lub średnie (9–12 mm.), następnie w fazie proliferacji i początkach przebudowy rośnie, co związane jest z przyrostem włókien ścięgnistych, by finalnie, w większości przypadków, ponownie zmaleć w wyniku przebudowy struktury i pozostawienia jedynie dobrze ukierunkowanych, wytrzymały struktur. Dla pacjentów 115805 i 127752, w ocenie radiologa, proces przebiega zgodnie z wyżej przedstawionym opisem. W przypadku pacjenta 140283 widoczna jest anomalia w 3 tyg., co może mieć związek z fluktuacjami np. nadmiernym obciążeniem ścięgna. Ponownie, jedynie w przypadku pacjenta 127108, gojenie przebiega odmiennie i po operacji pogrubienie rośnie, aż do 52 tyg.

Błąd (MAE) osiągnięty dla oceny tego parametru jest najniższy spośród wszystkich sześciu, co widoczne jest w przebiegach krzywych gojenia, które przybliżają ocenę radiologa z MAE równym 0.56. Anomalia z 3 tyg. występująca u pacjenta 140283 nie została dobrze wykryta i w rezultacie błąd maksymalny (MAX-AE) wyniósł 1.82. W celu detekcji przyczyny konieczna jest dalsza analiza i w pierwszej kolejności weryfikacja, czy błąd popełniony został przez ocenę radiologa, czy metodę.

Następnym analizowanym parametrem jest STE. Wykresy krzywych gojenia przedstawione na Rys. 5.15. Parametr ten związany jest z oceną granic pomiędzy ROI, a

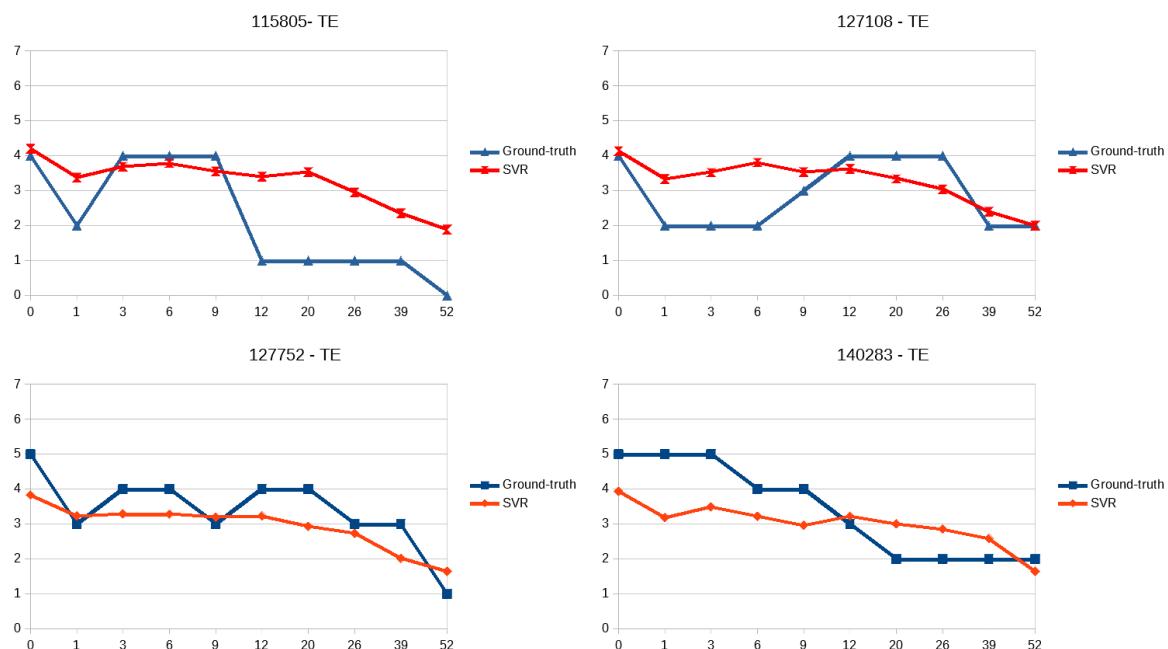


Rysunek 5.15: Ocena parametru STE.

zewnętrznymi tkankami. Ściegno po zerwaniu nigdy nie odzyskuje fraktałnych, zdro-

wych granic, stąd ocena radiologa oscylująca w zakresie 4–6. Zakres oceny automatycznej jest jeszcze mniejszy i wynosi po zaokrągleniu 4–5. Skutkiem tak niskich zakresów jest najgorszy wynik miary Corr spośród wszystkich sześciu parametrów, co oznacza, że drobne zmiany we fraktalności krawędzi nie są dobrze oceniane przez proponowaną metodę automatyczną. Z uwagi na mały zakres zmian, MAE i MAX-AE osiągnęły drugi najlepszy wynik. Na tym etapie informatywność tego parametru pozostaje zatem wątpliwa i konieczna jest dalsza weryfikacja.

Jako czwarty parametr przedstawiono TE. Uzyskane krzywe gojenia zaprezentowano na Rys. 5.16. W tym przypadku ocena radiologa skupiona jest na analizie płynów



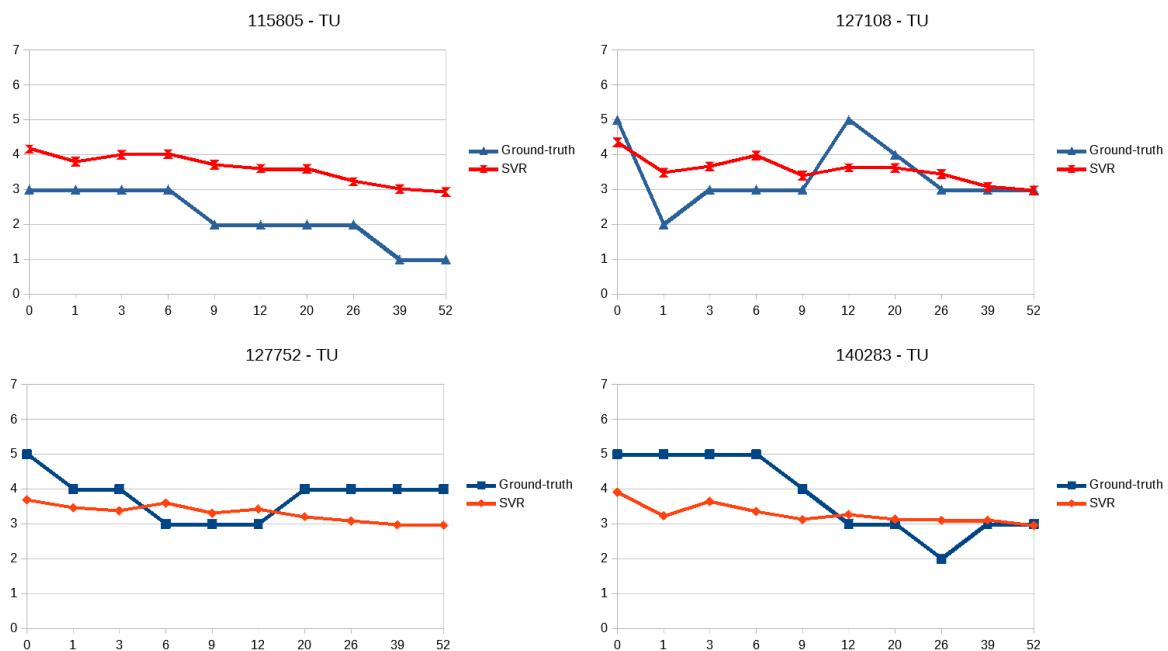
Rysunek 5.16: Ocena parametru TE.

gromadzących się w obszarze ROI, powodujących jego obrzęk. W większości przypadków, zszycie ścięgna powoduje redukcję płynów w ROI, która następnie w wyniku intensywnych procesów zapalnych i budowy nowych włókien ulega ponownej akumulacji. W ostatniej fazie przebudowy, z uwagi na zmniejszenie intensywności procesów, objętość płynów ponownie jest redukowana. Dla wszystkich czterech pacjentów ocena radiologa oddaje powyższy opis. W przypadku pacjenta 127108 można zaobserwować dłuższe fazy redukcji i akumulacji płynu, co ponownie potwierdza odmienność procesu gojenia dla tego przypadku. Interesujący jest przypadek pacjenta 140283, gdzie w 1 tygodniu po operacji minimum lokalne widoczne jest tylko w wyniku oceny automatycznej. W tym przypadku rozdzielncość skali radiologa okazała się być zbyt mała.

5.3 Ocena procesu *Pojedział 5: Nowa metoda oceny procesu gojenia ścięgna Achillesa*

Szukając powodów błędów maksymalnych, ponownie można wskazać za powód mniejszy zakres zmian oceny automatycznej. W 12 tyg. oceny pacjenta 115805 występuje silna redukcja obrzęku. W ocenie automatycznej widoczne jest lokalne minimum, ale jego wartość odbiega od wzorca o 2.48, a w następnym kroku o 2.52 osiągając MAX-AE.

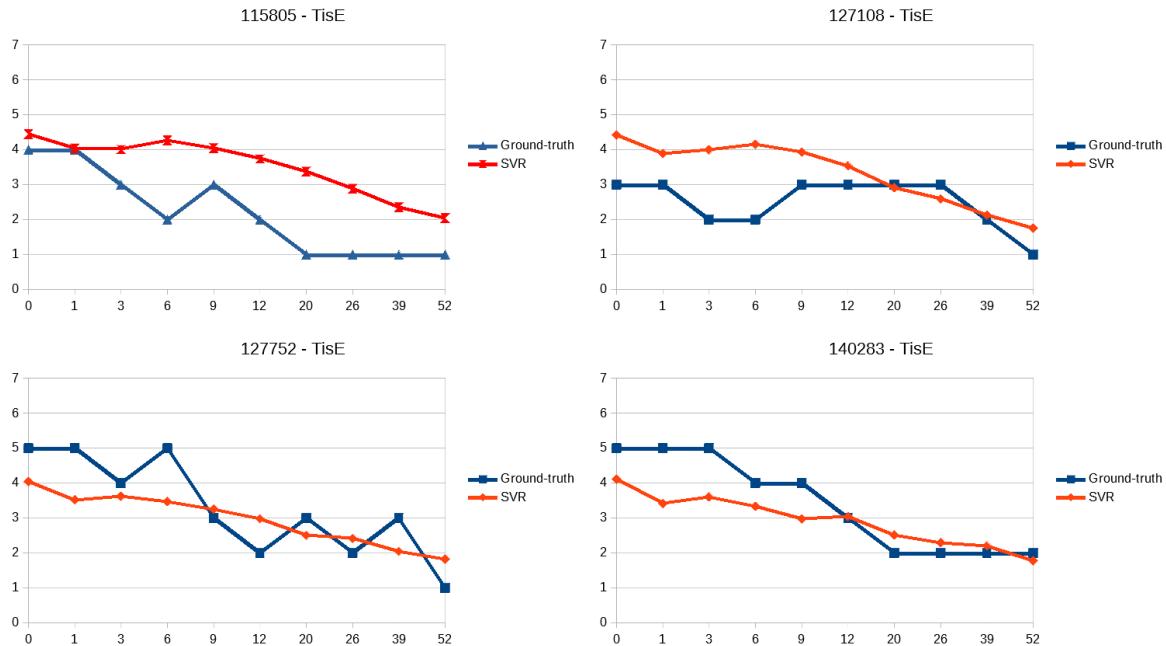
Przedostatnim z analizowanych parametrów jest TU. Krzywe gojenia zostały zaprezentowane na Rys. 5.17. Jest to jedyny parametr oceniany w całości w płaszczyźnie



Rysunek 5.17: Ocena parametru TU.

strzałkowej. Przy jego użyciu interpretowana jest jednorodność ścięgna widoczna w postaci charakterystycznego zdrowego warkocza włókien. Struktura ta w procesie gojenia się ulega stopniowej poprawie, co odzwierciedla oceny radiologa dla pacjentów 115805 i 140283 (z jedną anomalią w 26-tym tyg.). Dla pacjenta 127752 jednorodność ścięgna ulega pogorszeniu na przełomie 12 i 20 tyg., co może być wynikiem czynników zewnetrznych wpływających na przebieg gojenia. Natomiast krzywa dla pacjenta 127108 ponownie zawiera odmienne trendy, zwłaszcza na przełomie 9 i 12 tyg. Dla automatycznej oceny, parametr oceniany jest z drugim (zaraz po STE) najgorszym wynikiem pod względem Corr i również z drugim najgorszym MAE. Analiza automatyczna bazuje na metodzie wytrenowanej jedynie z wykorzystaniem przekrojów osiowych. Uzyskane wyniki są zatem efektem agregowanych w mierze H wartości lub korelacji z innymi parametrami, co utrudnia skuteczną charakterystykę zmian.

Jako ostatni został przeanalizowany parametr TisE. Charakteryzuje on obrzęk w przestrzeni powięziowej. Krzywe gojenia widoczne są na Rys. 5.18. Jest to jedyny pa-



Rysunek 5.18: Ocena parametru TisE.

rametr oceniany w całości na zewnątrz ROI tylko z wykorzystaniem cech DL. Na przełomie fazy poliferacji i remodelingu (6–9 tydz.) następuje charakterystyczny, krótkotrwały wzrost płynów w tym obszarze wynikający z intensywności procesu. W ocenach radiologa, szczególnie widoczny jest on dla pacjenta 115805 i 127752. Dla pacjenta 140283 widoczna jest stagnacja, a dla 127108 zwiększenie obrzęku, które utrzymuje się przez kolejne etapy do 26 tyg. W krzywych wygenerowanych w ocenie automatycznej widoczne są również maksima lokalne w okolicach przełomu faz gojenia się. Są one jednak przesunięte względem oceny radiologa (3–6 tydz.), co wymaga dalszej analizy.

Rozdział 6

Testy nowej metody

W tym rozdziale zostaną przedstawione porównania opracowanej metody. W pierwszej kolejności zaprezentowane zostanie zestawienie z metodą wytrenowaną z wykorzystaniem paradygmatu end-to-end, a zatem explicite modelującą ocenę radiologa. Następnie przedstawione zostanie porównanie z wynikami prac algorytmu oceniającego proces gojenia się ścięgna Achillea z wykorzystaniem danych z Ultrasonografii. Ostatnie zestawienie dotyczy porównania wyników z oceną biomechaniczną.

Zarówno metoda oparta o paradygmat end-to-end jak i o dane Ultrasonograficzne są wynikiem prac, w których autor tej rozprawy był współautorem, nadzorował grupę studentów opracowujących finalne rozwiązania oraz tworzył plan badań. W przypadku oceny biomechanicznej prace nad metodą zrealizowane zostały w całości przez mgr Magdalenę Syrek z Carolina Medical Center w Warszawie, zatem wartością dodaną tej rozprawy jest przedstawione porównanie.

6.1 Porównanie z metodą opartą o paradygmat end-to-end

W tej sekcji opracowana metoda zostanie porównana z metodą opartą o paradygmat end-to-end, explicite modelującą ocenę radiologa. Do tego zadania zostały wykorzystane trzy architektury sieci konwolucyjnych opisane w Rozdziale 4 tj. AlexNet, GoogLeNet (Inceptionv3) i ResNet z 50 warstwami konwolucyjnymi (ResNet-50). Wszystkie trzy architektury zostały zmodyfikowane poprzez zastąpienie końcowej warstwy klasyfikacyjnej warstwą regresji z pojedynczym wyjściem i funkcją kosztu zdefiniowaną jako

średni błąd kwadratowy. W skutek tego podejścia algorytm szkolony jest, aby minimizować błąd pomiędzy predykcją sieci neuronowej, a wartościami parametrów z wzorca odniesienia.

Do szkolenia się sieci wykorzystano powiększony zestaw treningowy, ograniczony jednak do danych z 44 pacjentów treningowych i trzech sekwencji wybranych w toku eksperymentów przedstawionych w podsekcji 5.2.3: T2 * GRE, T2 * GRE TE_MIN i PD. Tym razem nie zdecydowano się na zawężenie zbioru danych tylko do jednej sekwencji RM z uwagi na potrzebę maksymalizację liczebności zbioru. Nie zdecydowano się również na pełny zbiór wszystkich sekwencji, z uwagi na różnice w danych, które mogłyby zaburzać proces szkolenia się.

Do szkolenia zastosowano metodę kroswalidacji z podziałem na 4 segmenty. Dla każdego z parametrów z wzorca odniesienia wyszkolono osobną sieć. Wyliczono metryki MAE, MAX-AE i Corr stosowane również w poprzednim rozdziale. Wyniki zestawiono w Tab. 6.1.

Tabela 6.1: Wyniki szkolenia się sieci w paradygmacie end-to-end z wykorzystaniem zbioru treningowego.

		SCT	TT	STE	TE	TU	TisE
AlexNet_e	MAE	1.04±0.33	0.82±0.27	0.95±0.48	0.88±0.38	1.03±0.36	0.78±0.26
	MAX-AE	1.91	1.67	2.39	2.20	1.90	1.35
	Corr	0.82	0.72	0.13	0.70	0.50	0.83
Inceptionv3_e	MAE	0.88±0.32	0.75±0.22	0.82±0.33	0.78±0.20	0.91±0.34	0.67±0.23
	MAX-AE	1.57	1.44	1.93	1.17	1.85	1.11
	Corr	0.85	0.72	0.05	0.72	0.52	0.84
ResNet-50_e	MAE	0.64±0.21	0.98±0.28	0.83±0.37	0.89±0.22	1.06±0.41	1.10±0.34
	MAX-AE	1.21	1.79	1.94	1.43	2.42	2.36
	Corr	0.06	0.21	0.00	0.11	0.14	0.11

Otrzymane wartości MAE znajdują się w zakresie 0.64 do 1.10 (w skali wzorca odniesienia 0–7). Dla najlepszego modelu pod względem średniej MAE tj. Inceptionv3, średni dystans między predykcją i wzorcem wyniósł < 0.92. Najlepiej estymowanymi parametrami były TisE – 0.67, TT – 0.75 i TE – 0.78. Najgorsze rezultaty natomiast otrzymano dla TU – 0.91, który to parametr silnie zależy od informacji widocznej w płaszczyźnie strzałkowej.

W kolejnej tabeli (Tab. 6.2) przedstawiono zestawienie automatycznej oceny dla zbioru pacjentów testowych, realizowanej przez najlepszy model szkolony w paradygmacie end-to-end (InceptionV3) i proponowaną przez autora tej pracy metodą (SVR).

Metoda SVR osiąga wyższe rezultaty Corr w każdym z parametrów i niższe MAE dla 4 z 6 parametrów. Natomiast tylko w jednym z parametrów (tj. STE) charakteryzuje

Tabela 6.2: Test set results for the tendon healing progression.

Model		SCT	TT	STE	TE	TU	TisE
InceptionV3 _e	MAE	1.12±0.74	0.80±0.24	1.40±0.52	0.89±0.31	1.08±0.26	0.69±0.07
	MAX-AE	2.14	1.01	2.13	1.18	1.44	0.78
	Corr	0.82	0.77	0.05	0.59	0.02	0.77
SVR	MAE	1.05 ± 0.12	0.56 ± 0.06	0.75 ± 0.08	0.91 ± 0.10	0.91 ± 0.09	0.94 ± 0.10
	MAX-AE	2.62	1.82	1.92	2.54	2.01	2.38
	Corr	0.85	0.85	0.31	0.72	0.65	0.80

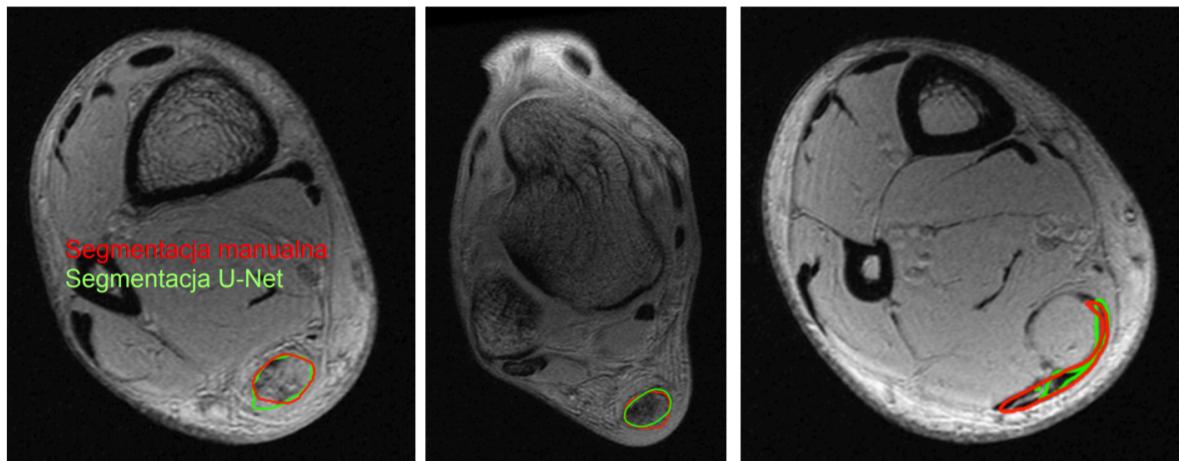
się niższym MAX-AE.

Można zatem wnioskować, że cechy wykorzystywane w metodzie SVR do oceny procesu gojenia się skuteczniej dywersyfikują kolejne fazy procesu i pozwalają na oczenną trendu z lepszą dokładnością niż metoda oparta o paradygmat end-to-end. 2 z 6 parametrów, w których pod względem MAE sieć InceptionV3 okazała się lepsza tj. TE i TisE dotyczą obrzęków, zatem elementów ocenianych pod kątem pola powierzchni i kształtu. Fakt ten należy interpretować, iż tego rodzaju informacja dominuje w ekstrahowanych przez jądra splotu obrazach, co z kolei implikuje gorsze wyniki w pozostałych parametrach ocenianych na podstawie wzorców tworzonych przez struktury ścięgniste. Wynik MAX-AE nie jest zaskakujący, ponieważ z uwagi na naturę treningu (explicite modelowanie wzorca odniesienia) maksymalne błędy będą minimalizowane. Trudność w interpretacji polega na fakcie, że maksymalne błędy mogą być wynikiem błędu algorytmu, jak również pojedynczymi omyłkami doświadczonego radiologa, który z uwagi np. na zmęczenie bądź inne dystrukcje zaburzające percepcję, popełnił błąd w ocenie.

W tym kontekście należy również podkreślić zasadnicze różnice między dwoma przedstawionymi metodami. Do rozwoju metody end-to-end niezbędna jest pełna informacja od radiologa tj. ocena 6 parametrów w kolejnych krokach czasowych. Średnio jest to około 1 godziny pracy na pacjenta. Proponowana przez autora metoda wykorzystuje sieć wyszkoloną na bazie etykiet wyższego poziomu tj. rozróżniających jedynie oznaczenia zdrowych i chorych tkanek. Takie szkolenie jest nazywane słabobadzorowanym (w odróżnieniu od pełni-badzorowanego w paradygmacie end-to-end). Dalszy rozwój proponowanej metody można zatem uzyskać przy wykorzystaniu danych oznaczonych jako chory, zdrowy oraz znacząco mniejszego zbioru pacjentów z oznaczonym pełnym wzorcem odniesienia, który posłużyłby do sporadycznych ulepszeń etapu meta-regresji.

Pewnym minusem proponowanego podejścia jest wymóg informacji o ROI, jednakże jego oznaczenie to czas około 10 minut na pacjenta, dla radiologa wykorzystującego odpowiednie oprogramowanie np. Osirix [X] lub VisNow [X]. Ponadto segmentację ROI

można stosunkowo łatwo zautomatyzować. Wstępne prace dr. Jędrzeja Nowosielskiego i dr. Piotra Regulskiego oraz autora tej pracy pokazują, że skuteczne w tym zakresie mogą być sieci typu *fully convolutional neural networks*. Przykład działania takiej architektury (tj. U-Net zaproponowanej w [X]) można zaobserwować na Rys. 6.1.



Rysunek 6.1: Automatyczna segmentacja ROI z wykorzystaniem głębokich sieci neuronowych.

Kolorem czerwonym oznaczono obszar segmentacji manuowej, wykonanej przez eksperta radiologa. Kolorem zielonym efekt segmentacji automatycznej. Miara DICE (zob. [X]) dla otrzymanych obrazów wynosi około 0.75 i świadczy o wysokiej jakości segmentacji oraz o obiecującym kierunku tego rodzaju prac.

Ciekawym elementem rozwoju automatycznej oceny byłaby również fuzja obu metod tj. podmiana ekstraktora cech DL trenowana na binarnie oznaczonym zbiorze na ekstraktor uzyskanego modelu InceptionV3. Jednak z uwagi na omawiane problemy praktyczne z dalszym rozwojem tak utworzonej metody oraz brak obiecujących rezultatów w przeprowadzonych przez autora tej pracy badaniach wstępnych, taka propozycja nie została uwzględniona w prezentowanej rozprawie.

6.2 Porównanie z metodą opartą o dane z ultrasonografii

W tej sekcji opracowana metoda została porównana z automatyczną oceną bazującą na danych z Ultrasonografii. Ponownie, do utworzenia metody opartej o USG wykorzystano konwolucyjne sieci neuronowe, a dokładniej AlexNet, InceptionV3 i ResNet-50

oraz wykorzystano omówiony w poprzedniej sekcji paradygmat end-to-end.

Dane dla metody opartej o Ultrasonografię pochodząły również z projektu START. Stosowano się zatem do identycznych odstępów czasowych, co w przypadku badań RM, a akwizycji poddano tych samych pacjentów. Z przyczyn praktycznych zmniejszyła się jedynie grupa odniesienia, która w tym przypadku wyniosła 18-stu zdrowych ochotników. Badania zrealizowano z wykorzystaniem aparatu GE 3D high-resolution Voluson E8 Expert z liniową sondą (5–18 MHz). Jako dane wejściowe wykorzystano informacje z trybu B (zob. 3.3), których ostateczna liczba wyniosła 565 skanów 3D.

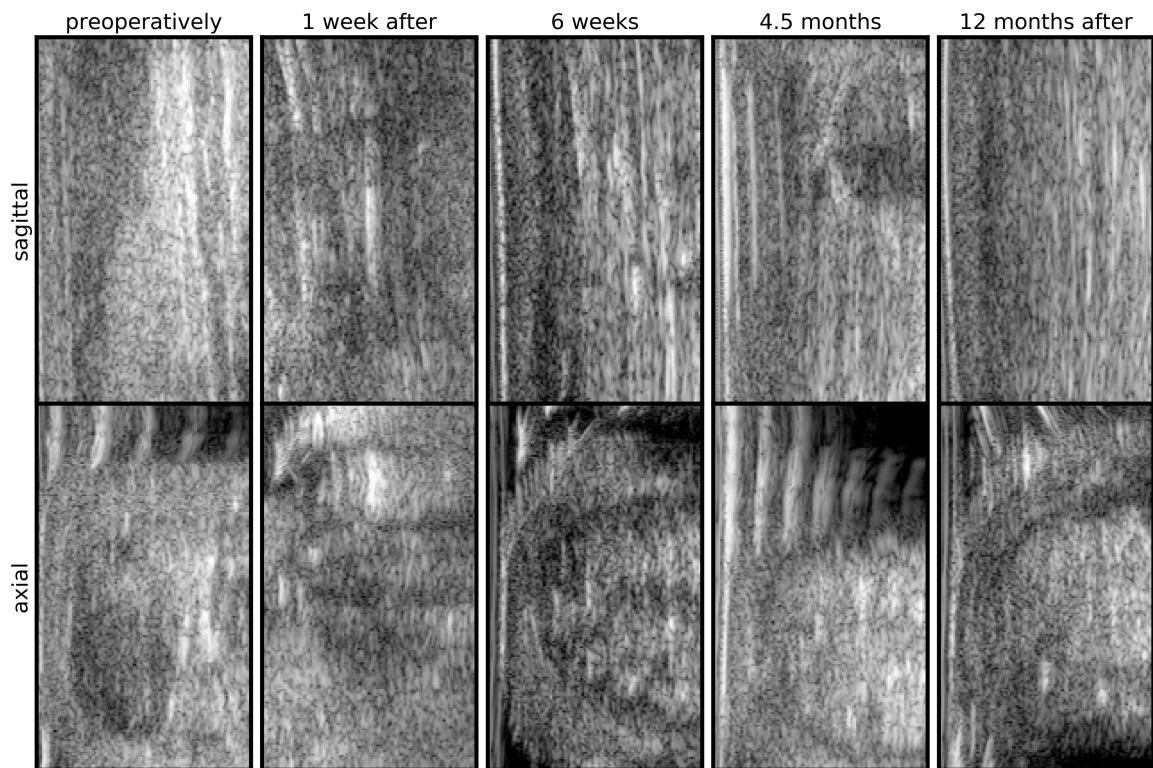
Dane USG są zbliżone do izotropowych dlatego utworzono zbiory zarówno w oparciu o przekroje w płaszczyźnie osiowej jak i strzałkowej (tylko tam, gdzie widoczne jest ścięgno):

- zbiór treningowy USG (strzałkowy) – zawiera 253.639 2D przekrojów w płaszczyźnie strzałkowej, w tym 245.366 pochodzących od chorych 44 pacjentów oznaczonych przez radiologa i 8.273 pochodzących od zdrowych ochotników.
- zbiór treningowy USG (osiowy) – zawiera 467.548 2D przekrojów w płaszczyźnie osiowej, w tym 450.816 pochodzących od chorych 44 pacjentów oznaczonych przez radiologa i 16.732 pochodzących od zdrowych ochotników.

Wizualizacja przykładowych danych USG znajduje się na Rys. 6.2. W ogólności można zaobserwować ułożenie włókien ścięgnistych na przekrojach strzałkowych i teksturę oraz tkanki otaczające ścięgno na przekrojach osiowych. Jednak szczegółowa analiza załączonych obrazów może być wykonana jedynie przez specjalistę radiologa.

Utrudniona w stosunku do obrazów RM interpretacja USG ma związek z obecnymi artefaktami w USG i z szumem ziarnistym tworzącym losowe wzorce. Jakość obrazu obniża również jego wartość kliniczną, co wskazują analizy zamieszczone w tej pracy jak i w wielu pracach innych autorów (zob. np. [63, 67]). Dlatego na początku zdecydowano się przeprowadzić test polegający na prostym zadaniu klasyfikacji binarnej chory/zdrowy (podobnie jak dla badań RM zamieszczonych w Sekcji 5.2.1). Ma to na celu porównanie możliwości interpretacji wnioskowania na podstawie obrazów RM i USG. Wyniki dla szkolenia z wykorzystaniem kroswalidacji z 4 segmentami zamieszczono w Tab. 6.3.

Wszystkie modele wyszkolono oddzielnie na przekrojach osiowych i oddzielnie na strzałkowych. Żeby zbalansować dane, przekroje od zdrowych ochotników powiększono



Rysunek 6.2: Wizualizacja przykładowych danych USG w kolejnych tygodniach po zszyciu ścięgna w przekrojach osiowych i strzałkowych.

poprzez odbicie, a przekroje od chorych pacjentów próbkowano. Dokładność dla najlepszego modelu w przekroju osiowym wyniosła 91,6%, a w przekroju strzałkowym 91,2%. Dla porównania, dokładność binarnej klasyfikacji na danych z RM wyniosła 99,83%. Wyniki te potwierdzają możliwość lepszej interpretacji danych RM dla zadanego problemu, jednak nie dyskryminują USG. W celu polepszenia powyższych rezultatów eksperymentowano również z możliwością klasyfikacji binarnej z wykorzystaniem jedynie obszaru ROI, jednakże działania te przyniosły odwrotny skutek.

Kolejnym krokiem eksperymentu było porównanie jakości oceny procesu gojenia.

Tabela 6.3: Wyniki szkolenia się sieci na zbiorze treningowym dla danych z USG.

Architecture	sagittal				axial		
	Accuracy	Precision	Recall		Accuracy	Precision	Recall
AlexNet	0.846±00.087	0.92±00.08	0.78±00.11		0.843±00.075	0.93±00.06	0.73±00.11
Inception-v3	0.916±00.049	0.97±00.04	0.90±00.06		0.901±00.052	0.95±00.3	0.87±00.07
ResNet-50	0.907±00.039	0.96±00.05	0.89±00.08		0.912±00.046	0.95±00.04	0.88±00.06

W badaniu wykorzystano ponownie paradygmat end-to-end i zmianę warstwy klasyfikującej na regresję z funkcją kosztu określona jako średni błąd kwadratowy. Dla tej metody próbowano również odtworzyć schemat sprawdzony dla RM, bazujący na treningu binarnym oraz ekstrakcji i redukcji cech, jednak w przypadku USG otrzymane wyniki nie były na tyle satysfakcyjne, aby zamieszczać je w tej pracy. Ostatecznie wyniki szkolenia sieci w paradygmacie end-to-end zamieszczono w Tab. 6.4.

Tabela 6.4: 5CV results for the tendon healing progress using end-to-end approach

Network		sagittal					
		SCT	TT	STE	TE	TU	TisE
AlexNet	MAE	0.96±0.41	0.80±0.27	0.82±0.29	0.95±0.41	0.87±0.39	1.08±0.47
	MAX-AE	1.75	1.32	1.87	1.35	1.61	2.03
	Corr	0.53±0.47	0.69±0.38	0.11±0.32	0.68±0.44	0.31±0.51	0.22±0.53
Inception-v3	MAE	0.88±0.35	0.67±0.23	0.80±0.31	0.82±0.23	0.84±0.32	0.93±0.29
	MAX-AE	1.69	1.32	1.69	1.31	1.58	1.64
	Corr	0.83±0.44	0.71±0.40	0.19±0.34	0.64±0.47	0.56±0.40	0.71±0.40
ResNet50	MAE	0.89±0.12	0.74±0.15	0.83±0.22	0.81±0.31	0.92±0.31	0.99±0.32
	MAX-AE	1.53	1.22	1.64	1.43	1.67	1.71
	Corr	0.62±0.31	0.38±0.51	0.23±0.41	0.62±0.51	0.12±0.43	0.43±0.50
axial							
		SCT	TT	STE	TE	TU	TisE
AlexNet	MAE	0.98±0.39	0.83±0.33	0.82±0.35	0.94±0.52	0.95±0.50	0.86±0.28
	MAX-AE	1.79	1.36	1.86	1.41	1.61	1.59
	Corr	0.45±0.33	0.62±0.39	0.20±0.45	0.60±0.47	0.03±0.42	0.59±0.40
Inception-v3	MAE	1.03±0.46	0.70±0.24	0.76±0.26	0.86±0.19	0.87±0.32	0.85±0.25
	MAX-AE	2.52	1.45	1.45	1.24	1.67	1.56
	Corr	0.77±0.47	0.69±0.40	0.22±0.37	0.65±0.41	0.55±0.44	0.72±0.41
ResNet50	MAE	1.05±0.31	0.78±0.33	0.80±0.24	1.02±0.25	0.87±0.15	0.91±0.28
	MAX-AE	1.98	1.51	1.59	1.63	1.45	1.57
	Corr	0.52±0.41	0.47±0.44	0.22±0.35	0.65±0.55	0.18±0.54	0.57±0.39

Najlepsze wyniki otrzymano dla sieci Inception-V3 i ResNet-50 dlatego te modele zostały dalej porównane z prezentowaną w tej pracy metodą. W tym celu wyniki MAE, MAX-AE i Corr zostały wyliczone dla zbioru pacjentów testowych i zebrane w Tab. 6.5.

Sieć, uzyskująca najlepsze wyniki, wytrenowana na USG tj. Inception-v3, uzyskała wartości MAE w zakresie 0,53 do 0,87 i średnioną korelację Z-Fisher od 0,31 do 0,8. Metoda oparta o USG charakteryzuje się lepszymi wynikami Corr tylko dla parametrów STE i TU, czyli dla parametrów ocenianych przez wzgląd na krawędzie i na osi w płaszczyźnie strzałkowej. W przypadku mierzenia MAE, metoda RM uzyskuje lepszy wynik w parametrze TT, gdzie znaczący wpływ na ocenę mają cechy z ROI.

Bazując na powyższych wynikach można wnioskować, że metoda oparta o dane USG może być komplementarna z proponowaną oceną bazującą na danych z RM. Zwłaszcza w kontekście oceny parametru TU jak i zmniejszenia błędu MAE w poszczególnych parametrach. Do tego celu potrzebna jest jednak dokładna analiza i zrozumienie podstaw wnioskowania sieci opartej o USG, a następnie wybór oraz realizacja

Tabela 6.5: Porównanie wyników oceny automatycznej bazującej na danych USG i RM na zbiorze testowym.

		sagittal					
Network		SCT	TT	STE	TE	TU	TisE
Inception-v3	MAE	0.81 \pm 0.38	0.63 \pm 0.06	0.56 \pm 0.18	0.85 \pm 0.20	0.54 \pm 0.04	0.87 \pm 0.29
	Corr	0.80 \pm 0.39	0.77 \pm 0.28	0.31 \pm 0.33	0.52 \pm 0.36	0.69 \pm 0.34	0.62 \pm 0.52
ResNet-50	MAE	0.88 \pm 0.33	0.65 \pm 0.15	0.66 \pm 0.09	0.83 \pm 0.25	0.75 \pm 0.12	0.93 \pm 0.22
	Corr	0.60 \pm 0.32	0.55 \pm 0.38	0.25 \pm 0.27	0.55 \pm 0.41	0.34 \pm 0.29	0.56 \pm 0.38
		axial					
		SCT	TT	STE	TE	TU	TisE
Inception-v3	MAE	0.84 \pm 0.54	0.75 \pm 1.45	0.58 \pm 0.10	0.83 \pm 0.10	0.53 \pm 0.16	0.83 \pm 0.30
	Corr	0.69 \pm 0.49	0.68 \pm 0.41	0.45 \pm 0.15	0.51 \pm 0.42	0.66 \pm 0.16	0.68 \pm 0.39
ResNet-50	MAE	0.92 \pm 0.37	0.76 \pm 0.32	0.68 \pm 0.08	0.81 \pm 0.17	0.65 \pm 0.20	0.94 \pm 0.11
	Corr	0.55 \pm 0.41	0.57 \pm 0.38	0.35 \pm 0.31	0.44 \pm 0.39	0.39 \pm 0.35	0.61 \pm 0.33
		RM					
		SCT	TT	STE	TE	TU	TisE
SVR	MAE	1.05 \pm 0.12	0.56 \pm 0.06	0.75 \pm 0.08	0.91 \pm 0.10	0.91 \pm 0.09	0.94 \pm 0.10
	MAX-AE	2.62	1.82	1.92	2.54	2.01	2.38
	Corr	0.85	0.85	0.31	0.72	0.65	0.80

odpowiedniej metody fuzji.

6.3 Porównanie z metodą opartą o badania biomechaniczne

W tej sekcji opracowana metoda została porównana z wynikami badań biomechanicznych i wynikami funkcjonalnego testu ATRS. Badania te były realizowane w ramach protokołu monitorowania rehabilitacji pacjentów opracowanego w klinice Carolina Medical Center (zob. Sekcja 3.4).

W ramach projektu START możliwe było zebranie danych dla 30 spośród 60 pacjentów. Dane zawierały wyniki pomiaru zrealizowanego w dwóch krokach czasowych tj. w 26 i 52 tygodniu. Po wstępnych analizach, do porównań w tej pracy wykorzystano zebrane wyniki ATRS i deficyty sił mięśniowych będące rezultatem badań biomechanicznych. Dokładnie, określono 16 takich deficytów oznaczonych dalej w tej pracy $d1-d16$. W czterech pozycjach tj. kolano zgięte i wyprostowane oraz staw skokowy zgięty wyprostowany zrealizowano pomiary w warunkach izometrii i izokinetyki w trzech prędkościach kątowych $60^\circ/s$, $120^\circ/s$ oraz $180^\circ/s$

Z uwagi na dużą liczbę pomiarów, w pierwszej kolejności zdecydowano się na przeprowadzenie analizy czynników głównych i identyfikację nowych zmiennych wyjaśniających

w największym stopniu wariację w zebranym zbiorze danych. Wyniki podsumowano w Tabeli 6.6.

Tabela 6.6: Wyniki analizy czynników głównych dla 16-tu pomiarów deficytów mięśniowych w warunkach izometrii i izokinetyki. Oznaczono ładunki większe niż 0,7.

Zmienna	Czynnik1	Czynnik2	Czynnik3	Czynnik4
d1	-0,77	0,16	-0,04	-0,25
d2	0,32	-0,74	0,038	-0,044
d3	-0,81	0,26	-0,17	-0,24
d4	0,09	-0,62	-0,50	-0,38
d5	-0,82	0,13	-0,03	0,21
d6	0,31	-0,53	-0,53	0,42
d7	-0,67	-0,20	-0,34	-0,11
d8	-0,12	-0,34	-0,67	0,43
d9	-0,61	-0,01	-0,28	-0,36
d10	0,07	-0,84	0,26	0,02
d11	-0,81	-0,14	0,01	0,08
d12	-0,11	-0,93	-0,01	-0,01
d13	-0,72	-0,24	0,28	0,36
d14	-0,12	-0,81	0,33	-0,25
d15	-0,58	-0,32	0,39	0,53
d16	-0,17	-0,62	0,23	-0,36
Udział	0,28	0,27	0,10	0,09

Z czterech czynników, 2 mają wyraźnie większy udział w wyjaśnianej wariancji, osiągając odpowiednio poziomy 28 i 27%. Istotne jest również, że 3 z pięciu znaczących ładunków czynnikowych ($d1$, $d3$, $d5$) dla pierwszego czynnika znajdują się w przedziale $d1-d8$, które skojarzone są z pozycją kolana prostego oraz, że dla drugiego czynnika głównego sytuacja jest odwrotna i 3 z czterech ładunków czynnikowych ($d10$, $d12$, $d14$) znajdują się w przedziale $d9-d16$, czyli badania przy kolanie zgiętym.

Co oczywiste, pozycja kolana wpływa znacząco na wynik badania biomechanicznego, dlatego w ramach dalszej pracy zdecydowano się na analizę czynnikową dla wyprostowanego i zgiętego kolana oddzielnie. Wyniki zostały przedstawione w Tabeli 6.7.

Ponownie można zaobserwować charakterystyczne rozróżnienie, kiedy to pierwszy czynnik główny dla badania w pozycji kolana wyprostowanego zawiera ładunki znaczące o nieparzystych numerach, a drugi czynnik posiada znaczące ładunki o parzystych

Tabela 6.7: Wyniki analizy czynników głównych dla 16-tu pomiarów deficytów mięśniowych, oddzielnie dla pozycji kolana zgiętego i wyprostowanego. Oznaczono ładunki większe niż 0,7.

Zmienna	Czynnik1	Czynnik2	Zmienna	Czynnik1	Czynnik2
d1	-0,82	-0,14	d9	-0,26	0,64
d2	-0,60	-0,37	d10	-0,72	-0,47
d3	0,88	-0,20	d11	-0,52	0,73
d4	-0,36	-0,65	d12	-0,83	-0,34
d5	0,83	-0,26	d13	-0,65	0,64
d6	-0,57	-0,64	d14	-0,81	-0,43
d7	0,55	-0,60	d15	-0,67	0,40
d8	-0,03	-0,80	d16	-0,65	-0,38
Udział	0,41	0,26		0,44	0,27

numerach. W przypadku badania w pozycji kolana zgiętego sytuacja jest odwrotna i analogicznie można zaobserwować rozróżnienie. W tym przypadku podział wynika z pozycji stawu skokowego. Ładunki o nieparzystych numerach posiadają czynniki związane z badaniami w zgięciu stawu skokowego, a o parzystych w wyproście. Z uwagi na ten fakt do kolejnych porównań zostały zdefiniowane następujące 4 zmienne:

1. ZZ – wyniki dla pozycji kolano zgięte i staw skokowy zgięty.
2. ZW – wyniki dla pozycji kolano zgięte i staw skokowy wyprost.
3. WZ – wyniki dla pozycji kolano wyprostowane i staw skokowy zgięty.
4. WW – wyniki dla pozycji kolano wyprostowane i staw skokowy wyprostowany.

Każda z nowo powstałych zmiennych składa się z czterech pomiarów: jednego dla izometrii i trzech dla izokinetyki. Dokładne zestawienie znajduje się w Tabeli 6.8.

Taki podział wydaje się być intuicyjny, ale fakt otrzymania powyższych wyników w analizie numerycznej jest wartością dodaną potwierdzającą spójność danych i jakość przeprowadzonych badań przez fizjoterapeutów z Carolina Medical Center.

W celu efektywnego porównania nowych zmiennych opisujących biomechanikę ścięgna i wyniku ATRS z predykcjami automatycznej metody oraz oceną radiologa, zdecydowano się dla badań radiologicznych przeprowadzić również analizę czynnikową. W obu przypadkach tj. oceny automatycznej i radiologa uzyskano minimalny udział w

Tabela 6.8: Komponenty i ich udział, wchodzące w skład nowych zmiennych opisujących biomechanikę.

Zmienna	Czynnik1	Czynnik2	Czynnik3	Czynnik4	Udział
ZZ	d9 (-0,61)	d11 (-0,88)	d13 (-0,92)	d15 (-0,77)	0,65
ZW	d10 (-0,86)	d12 (-0,91)	d14 (-0,92)	d16 (-0,77)	0,75
WZ	d1 (-0,85)	d3 (-0,90)	d5 (-0,87)	d7 (-0,72)	0,70
WW	d2 (-0,64)	d4 (-0,76)	d6 (-0,88)	d8 (-0,66)	0,55

wyjaśnianej wariancji czynników innych niż pierwszy. Stąd możliwość wprowadzenia nowej zmiennej tj. czynnika1 zamiast 6 parametrów: *SCT*, *STE*, *TE*, *TisE*, *TT*, *TU*. Szczegółowe wyniki znajdują się w Tabeli 6.9.

Tabela 6.9: Rezultaty analizy czynników głównych dla oceny radiologicznej i oceny automatycznej realizowanych z wykorzystaniem 6-ścio parametrycznej ankiety.
Oznaczono ładunki większe niż 0,7.

Zmienna	Czynnik1 - ocena radiologa	Czynnik1 - ocena automatyczna
SCT	-0,79	-0,95
STE	0,57	0,86
TE	-0,85	-0,94
TisE	-0,71	-0,89
TT	-0,76	-0,72
TU	-0,88	-0,94
Udział	0,57	0,78

Można zaobserwować, że udział wyjaśnianej wariancji w pierwszym czynniku jest większy dla oceny automatycznej niż dla oceny radiologa. Subiektywna ocena eksperta wprowadza nieuniknioną niepewność do pomiaru, stąd wyniki generowane przez człowieka są w większym stopniu losowe niż wyniki oceny automatycznej.

W kolejnych badaniach zestawione zostały interkorelacje nowych zmiennych. Na początku, w Tabeli 6.10, przedstawiono korelację oceny automatycznej z oceną radiologa.

Naturalnie, nowa zmienna opisująca ocenę radiologa koreluje z nową zmienną opisującą ocenę automatyczną. Fakt tej jest kolejnym potwierdzeniem na silny związek między zastosowanymi metodami przetwarzania obrazów i eksperckimi, ludzkimi spostrzeżeniami.

W kolejnej Tabeli (tj. Tabeli 6.11) przedstawiono wyniki korelacji nowych zmien-

Tabela 6.10: Korelacja między oceną radiologa, a oceną automatyczną. Oznaczone wsp. korelacji są istotne z $p < 0,05$ dla $N=30$.

	Ocena automatyczna	Ocena radiologa
Ocena automatyczna	1,0000	0,8070
Ocena radiologa	0,8070	1,0000

nych opisujących biomechanikę z wynikami testu ATRS.

Tabela 6.11: Korelacja badań biomechanicznych z testem ATRS. Oznaczone wsp. korelacji są istotne z $p < 0,05$ dla $N=30$.

	WW	WZ	ZW	ZZ	ATRS
WW	1,0000	-0,2037	0,4948	-0,0436	-0,2190
WZ	-0,2037	1,0000	-0,0143	0,5815	0,1070
ZW	0,4948	-0,0143	1,0000	0,2324	-0,2620
ZZ	-0,0436	0,5815	0,2324	1,0000	0,1437
ATRS	-0,2190	0,1070	-0,2620	0,1437	1,0000

Można zaobserwować istotną korelację pomiarów wykonanych przy takim samym ustawieniu kolana. Jest to istotniejsze nawet od ustawienia stawu skokowego. Z kolei analizując rezultaty korelacji z ATRS, można zaobserwować słabą korelację z wynikami pomiarów biomechanicznych. Uzasadnione jest zatem realizowanie w praktyce obydwu tych badań w celu skutecznego monitorowania rehabilitacji pacjentów.

W finalnym badaniu, w tej sekcji zestawiono nowe zmienne opisujące ocenę radiologa i ocenę automatyczną, z wynikami badań biomechanicznych i ATRS. Szczegółowe rezultaty przedstawiono w Tabeli 6.12.

Tabela 6.12: Korelacja badań biomechanicznych i testu ATRS z badaniami radiologicznymi ocenionymi automatycznie i przez radiologa. Oznaczone wsp. korelacji są istotne z $p < 0,05$ dla $N=30$.

	ocena automatyczna	ocena radiologa
WW	0,1197	0,0706
WZ	0,1576	0,1931
ZW	0,3429	0,3092
ZZ	0,3100	0,2996
ATRS	0,3854	0,2088

Można zaobserwować, że wskazania radiologa słabo korelują z wynikami badań bio-

*6.3 Porównanie z metodą opartą o badania biomecha*Rozdział 6: Testy nowej metody

mechanicznych oraz ATRS, co jest zgodne z praktyką kliniczną – dlatego zalecane jest wykonywanie badań radiologicznych, które dostarczając komplementarnych informacji.

Interesującym rezultatem jest wynik istotnej, niewielkiej korelacji oceny automatycznej z ATRS. Można wnioskować, że eliminacja losowości wynikającej z czynnika ludzkiego, spowodowała usystematyzowanie oceny, a jednocześnie uwspółnienie pewnych informacji, które obecne są również w ustrukturyzowanym teście ATRS. Wynik ten wyznacza ciekawy kierunek dalszych prac, celem odnalezienia informacji wspólnych i stworzenia komplementarnego testu maksymalizującego informację dla lekarza prowadzącego przy jednoczesnej minimalizacji objętości protokołu monitorowania pacjentów.

Rozdział 7

Podsumowanie

W ramach pracy udało się osiągnąć wszystkie postawione cele. Cel główny tj. opracowanie automatycznej metody oceny gojenia się ściegna Achillesa został osiągnięty poprzez zastosowanie unikalnej fuzji cech wyuczonych przez model głębokiej sieci neuronowej i cech wyekstrahowanych z wykorzystaniem klasycznych metod przetwarzania obrazów. Wnioski zostały spisane w Rozdziale 5 podzielonym na trzy sekcje. W pierwszej przedstawiono szczegółowy opis komponentów oraz ich parametrów wraz z unikalnymi w skali światowej danymi. W drugiej, zaprezentowano wyniki eksperymentów mających na celu dobrą komponentów i ich parametrów oraz optymalizacji czasu, jakości i kosztów związanych z zastosowaniem nowej metody. W ostatniej, trzeciej sekcji zaprezentowano szczegółową walidację wraz z analizą rozbieżności między nowym podejściem, a oceną radiologa.

W toku prac, zrealizowano również wszystkie cztery cele szczegółowe. Pierwszy, tj. "wybór efektywnego kosztowo i czasowo protokołu badania bazującego na technikach obrazowania medycznego, a dokładniej Rezonansu Magnetycznego" zrealizowano poprzez badania zamieszczone w p. 5.2.3. Najważniejszym rezultatem tych prac jest określenie utylitarnego protokołu składającego się tylko z jednej, trwającej jedynie 5 min. sekwencji RM.

Drugi cel, tj. "przetestowanie różnego rodzaju podejść związanych ze szkoleniem głębokich sieci neuronowych" zrealizowano poprzez zestawienie wyników działania nowej metody z wynikami alternatywnego podejścia tj. szkolenia sieci w oparciu o paradigmat end-to-end (zob. p. 6.1). W zestawieniu z alternatywną metodą analizy obrazów RM, wykazano przewagę zaproponowanego podejścia, a dokładniej uzyskano lepsze re-

zultaty korelacji w każdym z sześciu ocenianych parametrów oraz dla czterech z nich, niższy absolutny błąd średni.

Kolejny cel, tj. "Porównanie wyników oceny nowej metody z wynikami klasyfikacji bazującej na danych z Ultrasonografi" zrealizowano dołączając do pracy wyniki automatycznej interpretacji obrazów USG (zob. p. 6.2). W szczególności wykazano możliwość synergii obu podejść, bazując na większej izotropowości danych USG i dobrych wyników oceny parametrów widocznych w płaszczyźnie strzałkowej.

Finalnie, ostatni cel, tj. "porównanie wyników oceny nowej metody z oceną funkcjonalną, rutynowo stosowaną do wspomagania rehabilitacji po urazie ścięgna" osiągnięto uzyskując unikalne dane pomiarów biomechanicznych i testu ATRS, a następnie zestawiając je z wynikami nowo-opracowanej metody (zob. 6.3). Pokazano przydatność wykonywania badań radiologicznych, przez wzgląd na odmienność niesionej informacji, jednocześnie określono potencjalny obszar optymalizacji w zakresie niewielkiej, ale istotnej korelacji z wynikami testu ATRS.

W ramach powyższych prac nad nowym podejściem do oceny procesu gojenia się ścięgna Achillesa w badaniach RM, całość zadań dotyczących głębokich sieci neuronowych została zrealizowana przez autora tej pracy. Pełnił on również rolę głównego badacza, projektując i współrealizując prace nad fuzją cech (w szczególności w ramach testów perceptronu wielowarstwowego i SVR jako typów meta-regresji). Autor tej pracy nadzorował również i planował eksperymenty grupy tworzącej metodę bazującą na paradygmacie end-to-end oraz wykorzystywane do interpretacji obrazów USG, a następnie samodzielnie dokonał ich porównania z proponowanym w tej pracy podejściem.

Wszystkie wyniki mają charakter innowacyjny. Nie istnieje obecnie wedle najlepszej wiedzy autora tej pracy alternatywne podejście do usystematyzowanej i wyliczanej automatycznie oceny stanu gojenia się ścięgna Achillesa widocznego w badaniach RM. Rozprawa ta stanowi zatem kompleksowy opis obecnego *state-of-the-art* dla przedmiotowego problemu. Biorąc pod uwagę, że w przedstawionym stanie wyniki oceny automatycznej korelują z oceną radiologa nawet na poziomie 0,85, a minimalny uzyskany absolutny błąd średni równy jest 0,56, krytyczne w opinii autora tej pracy staje się polepszenie wzorca odniesienia. W dalszym toku prac planowana jest zatem stała współpraca m.in. z radiologami celem realizacji ocen przez kolejnych specjalistów oraz wielo-ośrodkowe zbieranie nowych wolumenów danych. Planowane jest również przeprowadzenie badań w zakresie usprawnienia przetwarzania danych, w szczególności w kwestii automatycznej segmentacji ROI. Finalnie, praca ta stanowi również

Rozdział 7: Podsumowanie

fundament pod rozpoczęcie nowych testów dotyczących pokrewnych zadań jak ocena gojenia się podobnych struktur typu więzadła stawu kolanowego, skokowego, barkowego, czy łokciowego. Polepszenie skuteczności wspomagania rehabilitacji w obszarze układu mięśniowo-szkieletowego stanowi według autora tej pracy ciekawe wyzwanie badawcze jak również ma wymiar praktyczny. Już teraz bowiem, według danych Krajowej Izby Fizjoterapeutów, po zwolnieniach z pracy dotyczących ciąży, zwolnienia na skutek dolegliwości mięśniowo-szkieletowych stanowią drugi powód absencji w pracy w Polsce.

Bibliografia

- [1] Adams msc software corp. Accessed: 2018-07-23.
- [2] Anybody tech. Accessed: 2018-07-23.
- [3] Bts systems. Accessed: 2018-07-23.
- [4] Deep learning: The next step in applied healthcare data. <https://insights.samsung.com/2016/07/12/deep-learning-the-next-step-in-applied-healthcare-data/>. Accessed: 2018-22-05.
- [5] Fizyka wobec wyzwań xxi wieku – obrazowanie medyczne. Accessed: 2018-07-30.
- [6] From challenges to impact of machine learning in clinical practice. <http://on-demand.gputechconf.com/gtc/2018/video/S8897/>. Accessed: 2018-22-05.
- [7] Ge healthcare. Accessed: 2018-06-08.
- [8] Kistler systems. Accessed: 2018-07-23.
- [9] Medical image net. <http://langlotzlab.stanford.edu/projects/medical-image-net/>. Accessed: 2018-22-05.
- [10] Meet iq. whole body imaging. under 2k. usd. Accessed: 2018-07-30.
- [11] Nvidia tensorrt. <https://developer.nvidia.com/tensorrt>. Accessed: 2018-22-05.
- [12] An overview of resnet and its variants. <https://towardsdatascience.com/an-overview-of-resnet-and-its-variants-5281e2f56035>. Accessed: 2018-22-05.

- [13] Project clara: Nvidia supercomputing platform redefines medical imaging. <https://blogs.nvidia.com/blog/2018/03/28/ai-healthcare-gtc/>. Accessed: 2018-07-30.
- [14] Queastions and answears in mri. Accessed: 2018-07-30.
- [15] Typy głowic stosowanych w usg. Accessed: 2018-06-08.
- [16] Us4us – ultrasound r&d in medicine, biology and industry. <http://us4us.eu/>. Accessed: 2018-07-30.
- [17] Vicon systems. Accessed: 2018-07-23.
- [18] Visual 3d - c motion. Accessed: 2018-07-23.
- [19] Salem Saleh Al-amri, Namdeo V. Kalyankar, Khamitkar S. D. Image segmentation by using threshold techniques. *CoRR*, abs/1005.4020, 2010.
- [20] Sanjeev Arora, Aditya Bhaskara, Rong Ge, Tengyu Ma. Provable bounds for learning some deep representations. *CoRR*, abs/1310.6343, 2013.
- [21] D. S. Babcock, H. Patriquin, M. LaFortune, M. Dauzat. Power doppler sonography: basic principles and clinical applications in children. *Pediatric Radiology*, 26(2):109–115, Feb 1996.
- [22] Tito Bassani, Elena Stucovitz, Zhihui Qian, Matteo Briguglio, Fabio Galbusera. Validation of the AnyBody full body musculoskeletal model in computing lumbar spine loads at l4l5 level. *Journal of Biomechanics*, 58:89–96, jun 2017.
- [23] Robert M. Bell, Yehuda Koren. Lessons from the netflix prize challenge. *SIGKDD Explor. Newslett.*, 9(2):75–79, dec 2007.
- [24] Richard Ernest Bellman. *Dynamic Programming*. Princeton University Press, 1957.
- [25] P. K. Beredjiklian. Biologic aspects of flexor tendon laceration and repair. *J Bone Joint Surg Am*, 85-A(3):539–550, Mar 2003.
- [26] F. Bloch. Nuclear induction. *Physical Review*, 70(7-8):460–474, oct 1946.
- [27] A. Bochenek, M. Reicher. *Anatomia człowieka*. Anatomia człowieka. PZWL, 2016.

- [28] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, Oct 2001.
- [29] G. Breit, I. I. Rabi. On the interpretation of present values of nuclear moments. *Physical Review*, 46(3):230–231, aug 1934.
- [30] Carolina Medical Center. Start - wykorzystanie autologicznych mezenchymalnych komórek macierzystych w procesie regeneracji rekonstruowanego ściegna achillesa, 2014.
- [31] Chun-houh Chen, Wolfgang Hrdle, Antony Unwin, Chun-houh Chen, Wolfgang Hrdle, Antony Unwin. *Handbook of Data Visualization (Springer Handbooks of Computational Statistics)*. Springer-Verlag TELOS, Santa Clara, CA, USA, wydanie 1, 2008.
- [32] Ilkay Zihni Chirali. 14 - sports injuries. Ilkay Zihni Chirali, redaktor, *Traditional Chinese Medicine Cupping Therapy (Third Edition)*, strony 211 – 230. Churchill Livingstone, wydanie third edition, 2014.
- [33] Junghwan Cho, Eunmi Lee, Hyunkwang Lee, Bob Liu, Xinhua Li, Shahein Tajmir, Dushyant Sahani, Synho Do. Machine learning powered automatic organ classification for patient specific organ dose estimation. *Society for Imaging Informatics in Medicine*, wolumen 2017, 2017.
- [34] Anna Choromanska, Mikael Henaff, Michaël Mathieu, Gérard Ben Arous, Yann LeCun. The loss surface of multilayer networks. *CoRR*, abs/1412.0233, 2014.
- [35] Eric M. Christiansen, Samuel J. Yang, D. Michael Ando, Ashkan Javaherian, Gaia Skibinski, Scott Lipnick, Elliot Mount, Alison O’Neil, Kevan Shah, Alicia K. Lee, Piyush Goyal, William Fedus, Ryan Poplin, Andre Esteva, Marc Berndl, Lee L. Rubin, Philip Nelson, Steven Finkbeiner. In silico labeling: Predicting fluorescent labels in unlabeled images. *Cell*, 173(3):792–803.e19, apr 2018.
- [36] Ryan Cunningham, Peter Harding, Ian Loram. Deep residual networks for quantification of muscle fiber orientation and curvature from ultrasound images. María Valdés Hernández, Víctor González-Castro, redaktorzy, *Medical Image Understanding and Analysis*, strony 63–73, Cham, 2017. Springer International Publishing.

- [37] Valentino Dardanoni, Giuseppe De Luca, Salvatore Modica, Franco Peracchi. Bayesian Model Averaging for Generalized Linear Models with Missing Covariates. EIEF Working Papers Series 1311, Einaudi Institute for Economics and Finance (EIEF), dec 2013.
- [38] Scott L. Delp, Frank C. Anderson, Allison S. Arnold, Peter Loan, Ayman Habib, Chand T. John, Eran Guendelman, Darryl G. Thelen. OpenSim: Open-source software to create and analyze dynamic simulations of movement. *IEEE Transactions on Biomedical Engineering*, 54(11):1940–1950, nov 2007.
- [39] S.L. Delp, J.P. Loan, M.G. Hoy, F.E. Zajac, E.L. Topp, J.M. Rosen. An interactive graphics-based model of the lower extremity to study orthopaedic surgical procedures. *IEEE Transactions on Biomedical Engineering*, 37(8):757–767, 1990.
- [40] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. *CVPR09*, 2009.
- [41] Denitsa Docheva, Sebastian A. Müller, Martin Majewski, Christopher H. Evans. Biologics for tendon repair. *Advanced Drug Delivery Reviews*, 84:222–239, apr 2015.
- [42] Mahmut Nedim Doral, Mahbub Alam, Murat Bozkurt, Egemen Turhan, Ozgür Ahmet Atay, Gürhan Dönmez, Nicola Maffulli. Functional anatomy of the achilles tendon. *Knee Surgery, Sports Traumatology, Arthroscopy*, 18(5):638–643, feb 2010.
- [43] Qi Dou, Hao Chen, Lequan Yu, Jing Qin, Pheng-Ann Heng. Multilevel contextual 3-d CNNs for false positive reduction in pulmonary nodule detection. *IEEE Transactions on Biomedical Engineering*, 64(7):1558–1567, jul 2017.
- [44] K. J. Dria, J. R. Sachleben, P. G. Hatcher. Solid-state carbon-13 nuclear magnetic resonance of humic acids at high magnetic field strengths. *J. Environ. Qual.*, 31(2):393–401, 2002.
- [45] John Duchi, Elad Hazan, Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. *J. Mach. Learn. Res.*, 12:2121–2159, jul 2011.

- [46] Bradley J. Erickson, Panagiotis Korfiatis, Zeynettin Akkus, Timothy Kline, Kenneth Philbrick. Toolkits and libraries for deep learning. *Journal of Digital Imaging*, 30(4):400–405, Aug 2017.
- [47] Bradley J. Erickson, Panagiotis Korfiatis, Timothy L. Kline, Zeynettin Akkus, Kenneth Philbrick, Alexander D. Weston. Deep learning in radiology: Does one size fit all? *Journal of the American College of Radiology*, 15(3):521–526, mar 2018.
- [48] Andre Esteva, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M. Swetter, Helen M. Blau, Sebastian Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639):115–118, jan 2017.
- [49] R. Frisch, O. Stern. Über die magnetische ablenkung von wasserstoffmolekülen und das magnetische moment des protons. i. *Zeitschrift für Physik*, 85(1):4–16, Jan 1933.
- [50] Ann Ganestam, Thomas Kallemose, Anders Troelsen, Kristoffer Weisskirchner Barfod. Increasing incidence of acute achilles tendon rupture and a noticeable decline in surgical treatment from 1994 to 2013. a nationwide registry study of 33, 160 patients. *Knee Surgery, Sports Traumatology, Arthroscopy*, 24(12):3730–3737, Luty 2015.
- [51] Matthew F. Glasser, Timothy S. Coalson, Emma C. Robinson, Carl D. Hacker, John Harwell, Essa Yacoub, Kamil Ugurbil, Jesper Andersson, Christian F. Beckmann, Mark Jenkinson, Stephen M. Smith, David C. Van Essen. A multi-modal parcellation of human cerebral cortex. *Nature*, 536(7615):171–178, jul 2016.
- [52] G. T. Gomes, S. Van Cauter, M. De Beule, L. Vigneron, C. Pattyn, E. A. Aude naert. *Patient-Specific Modelling in Orthopedics: From Image to Surgery*, strony 109–129. Springer Netherlands, Dordrecht, 2013.
- [53] Ian Goodfellow, Yoshua Bengio, Aaron Courville. *Deep Learning*. MIT Press, 2016.
- [54] Varun Gulshan, Lily Peng, Marc Coram, Martin C. Stumpe, Derek Wu, Arunachalam Narayanaswamy, Subhashini Venugopalan, Kasumi Widner, Tom Ma dams, Jorge Cuadros, Ramasamy Kim, Rajiv Raman, Philip C. Nelson, Jessica L.

- Mega, Dale R. Webster. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*, 316(22):2402, dec 2016.
- [55] Samuel R. Hamner, Ajay Seth, Scott L. Delp. Muscle contributions to propulsion and support during running. *Journal of Biomechanics*, 43(14):2709–2716, oct 2010.
- [56] R. Haralick, K. Shanmugam, I. Dinstein. Texture features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 3(6), 1973.
- [57] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.
- [58] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Identity mappings in deep residual networks. *CoRR*, abs/1603.05027, 2016.
- [59] Omar W. Heyward, Lucas M. Rabello, Lucas van der Woude, Inge van den Akker-Scheek, Alli Gokeler, Henk van der Worp, Johannes Zwerver. The effect of load on achilles tendon structure in novice runners. *Journal of Science and Medicine in Sport*, 21(7):661–665, jul 2018.
- [60] A. V. Hill. The heat of shortening and the dynamic constants of muscle. *Proceedings of the Royal Society B: Biological Sciences*, 126(843):136–195, oct 1938.
- [61] F.S. Hill, Stephen M. Kelly. *Computer Graphics using OpenGL*. Prentice Hall, 2006.
- [62] Geoffrey E. Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, Ruslan Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. *CoRR*, abs/1207.0580, 2012.
- [63] Noha Mohamed AbdelMaboud Ibrahim, Hytham Haroun Elsaeed. Lesions of the achilles tendon: Evaluation with ultrasonography and magnetic resonance imaging. *The Egyptian Journal of Radiology and Nuclear Medicine*, 44(3):581–587, sep 2013.
- [64] Jacques, Pierre Curie. Development, via compression, of electric polarization in hemihedral crystals with inclined faces. *Bulletin de la Societe de Minerologie de France*, (3):90–93, 1880.

- [65] Chand T. John, Frank C. Anderson, Jill S. Higginson, Scott L. Delp. Stabilisation of walking by intrinsic muscle properties revealed in a three-dimensional muscle-driven simulation. *Computer Methods in Biomechanics and Biomedical Engineering*, 16(4):451–462, apr 2013.
- [66] C. Ju, A. Bibaut, M. J. van der Laan. The Relative Performance of Ensemble Methods with Deep Convolutional Neural Networks for Image Classification. *ArXiv e-prints*, apr 2017.
- [67] K M Khan. Are ultrasound and magnetic resonance imaging of value in assessment of achilles tendon disorders? a two year prospective study. *British Journal of Sports Medicine*, 37(2):149–153, apr 2003.
- [68] Diederik P. Kingma, Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014.
- [69] Patrick Koch, Brett Wujek, Oleg Golovidov, Steven Gardner. Automated hyperparameter tuning for effective machine learning. 2017.
- [70] T Koivunen-Niemelä, K Parkkola. Anatomy of the achilles tendon (tendo calcaneus) with respect to tendon thickness measurements. *Surgical and Radiologic Anatomy*, 17(3):263–268, sep 1995.
- [71] Stanisław Kowalik. Zastosowanie teorii zbiorów przybliżonych do podejmowania decyzji. *Zarządzanie przedsiębiorstwem w erze postindustrialnej – ekonomia, prawo, kultura, etyka*, strony 39–44, Kazimierz Dolny, Polska, 2003.
- [72] Alex Krizhevsky. Learning multiple layers of features from tiny images. Raport instytutowy, 2009. Accessed: 2018-05-21.
- [73] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, NIPS’12, strony 1097–1105, USA, 2012. Curran Associates Inc.
- [74] Y. Kwon, M. Rhu. A case for memory-centric hpc system architecture for training deep neural networks. *IEEE Computer Architecture Letters*, 17(2):134–138, July 2018.

- [75] P. C. LAUTERBUR. Image formation by induced local interactions: Examples employing nuclear magnetic resonance. *Nature*, 242(5394):190–191, mar 1973.
- [76] K. N. Leach. A survey paper on independent component analysis. *Proceedings of the Thirty-Fourth Southeastern Symposium on System Theory (Cat. No.02EX540)*, strony 239–242, 2002.
- [77] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
- [78] Yann LeCun, Bernhard E. Boser, John S. Denker, Donnie Henderson, R. E. Howard, Wayne E. Hubbard, Lawrence D. Jackel. Handwritten digit recognition with a back-propagation network. D. S. Touretzky, redaktor, *Advances in Neural Information Processing Systems 2*, strony 396–404. Morgan-Kaufmann, 1990.
- [79] D.R. Lide. *CRC Handbook of Chemistry and Physics, 87th Edition*. Taylor & Francis, 2006.
- [80] Tony W. Lin, Luis Cardenas, Louis J. Soslowsky. Biomechanics of tendon injury and repair. *Journal of Biomechanics*, 37(6):865–877, jun 2004.
- [81] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, C. Lawrence Zitnick. Microsoft coco: Common objects in context. David Fleet, Tomas Pajdla, Bernt Schiele, Tinne Tuytelaars, redaktorzy, *Computer Vision – ECCV 2014*, strony 740–755, Cham, 2014. Springer International Publishing.
- [82] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A.W.M. van der Laak, Bram van Ginneken, Clara I. Sánchez. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60–88, dec 2017.
- [83] Mengchen Liu, Jiaxin Shi, Zhen Li, Chongxuan Li, Jun Zhu, Shixia Liu. Towards better analysis of deep convolutional neural networks. *CoRR*, abs/1604.07043, 2016.
- [84] Nicola Maffulli, Per Renström, Wayne B. Leadbetter, redaktorzy. *Tendon Injuries*. Springer-Verlag, 2005.

- [85] R. Makarewicz. *Kształtowanie pola akustycznego z uwzględnieniem aspektu akustyki środowiska*. Akustyka - Uniwersytet im. Adama Mickiewicza w Poznaniu. Wydawnictwo Naukowe Uniwersytetu im. Adama Mickiewicza, 1978.
- [86] P Mansfield. Multi-planar image formation using nmr spin echoes. *Journal of Physics C: Solid State Physics*, 10(3):L55, 1977.
- [87] J. Maquirriain. Achilles tendon rupture: avoiding tendon lengthening during surgical repair and rehabilitation. *Yale J Biol Med*, 84(3):289–300, Sep 2011.
- [88] Gary Marcus. Deep learning: A critical appraisal. *CoRR*, abs/1801.00631, 2018.
- [89] Stefano Markidis, Steven Wei Der Chien, Erwin Laure, Ivy Bo Peng, Jeffrey S. Vetter. NVIDIA tensor core programmability, performance & precision. *CoRR*, abs/1803.04014, 2018.
- [90] Miriam Martínez, Luis Enrique Sucar, Hector Gabriel Acosta, Nicandro Cruz. Bayesian model combination and its application to cervical cancer detection. Jaime Simão Sichman, Helder Coelho, Solange Oliveira Rezende, redaktorzy, *Advances in Artificial Intelligence - IBERAMIA-SBIA 2006*, strony 622–631, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.
- [91] Seymour Papert Marvin Lee Minsky. *Perceptrons: An Introduction to Computational Geometry*. Mit Press, 1969.
- [92] Warren S. McCulloch, Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, Dec 1943.
- [93] Ümit V. Çatalyürek, Cevdet Aykanat, Bora Uçar. On two-dimensional sparse matrix partitioning: Models, methods, and a recipe. *SIAM Journal on Scientific Computing*, 32(2):656–683, 2010.
- [94] Azim Celik Muhammed Elmaoglu. *Rezonans Magnetyczny: Podstawy fizyczne, obrazowanie, ułożenie pacjenta, protokoły*. Springer, 2015.
- [95] Kapinski N., Zielinski J., Borucki B., Trzcinski T., Ciszkowska Łyson B., Zdanowicz U., Smigielski R., Nowinski K. Monitoring of the achilles tendon healing process: can artificial intelligence be helpful? *ACTA OF BIOENGINEERING AND BIOMECHANICS*, 21(1), 2019.

- [96] Katarina Nilsson-Helander, Roland Thomeé, Karin Grävare-Silbernagel, Pia Thomeé, Eva Faxén, Bengt I. Eriksson, Jon Karlsson. The achilles tendon total rupture score (ATRS). *The American Journal of Sports Medicine*, 35(3):421–426, mar 2007.
- [97] Jędrzej Nowosielski, Jakub Zielinski, Bartosz Borucki, Krzysztof Nowinski. Multidimensional haralick's feature space analysis for assessment of the achilles tendon in mr imaging. *International Journal of Computer Assisted Radiology and Surgery*, 12(Supplement 1):218—220, 2017.
- [98] Karl F. Orishimo, Gideon Burstein, Michael J. Mullaney, Ian J. Krementz, Marcus Nesse, Malachy P. McHugh, Steven J. Lee. Effect of knee flexion angle on achilles tendon force and ankle joint plantarflexion moment during passive dorsiflexion. *The Journal of Foot and Ankle Surgery*, 47(1):34–39, jan 2008.
- [99] K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2:559–572, 1901.
- [100] Eric J. Perreault, Charles J. Heckman, Thomas G. Sandercock. Hill muscle model errors during movement are greatest within the physiologically relevant range of motor unit firing rates. *Journal of Biomechanics*, 36(2):211–218, Luty 2003.
- [101] M. S. Raikin. Epidemiology of achilles tendon rupture in the us, 2014.
- [102] P. Rajpurkar, A. Y. Hannun, M. Haghpanahi, C. Bourn, A. Y. Ng. Cardiologist-Level Arrhythmia Detection with Convolutional Neural Networks. *ArXiv e-prints*, jul 2017.
- [103] Piotr Regulski, Jakub Zieliński, Bartosz Borucki, Krzysztof Nowiński. Comparison of noise reducing t2map reconstruction methods in mri imaging of achilles tendon. *International Journal of Computer Assisted Radiology and Surgery*, 12(Supplement 1):15–16, Jun 2017.
- [104] Olaf Ronneberger, Philipp Fischer, Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015.
- [105] Frank Rosenblatt. The perceptron: A perceiving and recognizing automaton. Raport instytutowy, 1957.

- [106] A. Rosińska, M. Syrek. Rehabilitacja ściegna achillesa - kopletny protokół usprawniania po trójpęczkowej rekonstrukcji ściegna achillesa., 2017.
- [107] David E. Rumelhart, Geoffrey E. Hinton, Ronald J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, oct 1986.
- [108] N. Russell, T. Brendan Murphy, A. E Raftery. Bayesian model averaging in model-based clustering and density estimation. *ArXiv e-prints*, jun 2015.
- [109] Hijazi S., Kumar R., Rowen C. Using convolutional neural networks for image recognition. Raport instytutowy, Cadence, 2015. Accessed: 2018-05-21.
- [110] Saman Sarraf, Ghassem Tofighi. Deep learning-based pipeline to recognize alzheimer disease using fMRI data. *2016 Future Technologies Conference (FTC)*. IEEE, dec 2016.
- [111] W. Sawicki. *Histologia*. Wydawnictwo Lekarskie PZWL, 2008.
- [112] Jürgen Schmidhuber. Deep learning in neural networks: An overview. *Neural Networks*, 61:85 – 117, 2015.
- [113] Ramprasaath R. Selvaraju, Abhishek Das, Ramakrishna Vedantam, Michael Cogswell, Devi Parikh, Dhruv Batra. Grad-cam: Why did you say that? visual explanations from deep networks via gradient-based localization. *CoRR*, abs/1610.02391, 2016.
- [114] Dipanjan Sengupta, Yida Wang, Narayanan Sundaram, Theodore L. Willke. High-performance incremental svm learning on intel xeon phi processors. Julian M. Kunkel, Rio Yokota, Pavan Balaji, David Keyes, redaktorzy, *High Performance Computing*, strony 120–138, Cham, 2017. Springer International Publishing.
- [115] Qi Shao, Daniel N. Bassett, Kurt Manal, Thomas S. Buchanan. An EMG-driven model to estimate muscle forces and joint moments in stroke patients. *Computers in Biology and Medicine*, 39(12):1083–1088, dec 2009.
- [116] P. Sharma, N. Maffulli. Biology of tendon injury: healing, modeling and remodeling. *J Musculoskeletal Neuronal Interact*, 6(2):181–190, 2006.

- [117] Karen Simonyan, Andrea Vedaldi, Andrew Zisserman. Deep inside convolutional networks: Visualising image classification models and saliency maps. *CoRR*, abs/1312.6034, 2013.
- [118] Karen Simonyan, Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014.
- [119] Leon Sixt, Benjamin Wild, Tim Landgraf. Rendergan: Generating realistic labeled data. *CoRR*, abs/1611.01331, 2016.
- [120] Fengguang Song, Jack Dongarra. Scaling up matrix computations on shared-memory manycore systems with 1000 cpu cores. *Proceedings of the 28th ACM International Conference on Supercomputing*, ICS '14, strony 333–342, New York, NY, USA, 2014. ACM.
- [121] Katherine M. Steele, Ajay Seth, Jennifer L. Hicks, Michael S. Schwartz, Scott L. Delp. Muscle contributions to support and progression during single-limb stance in crouch gait. *Journal of Biomechanics*, 43(11):2099–2105, aug 2010.
- [122] Katarzyna Strojek, Zuzanna Piekorz, Urszula Kaźmierczak, Agnieszka Strączyńska, Walery Zukow. Etiologia i patomechanizm uszkodzenia Ściegna achillesa = the etiology and patomechanism damage to the achilles tendon. 2016.
- [123] R. M. Summers, P. M. Joseph, H. L. Kundel. Sodium nuclear magnetic resonance imaging of neuroblastoma in the nude mouse. *Invest Radiol*, 26(3):233–241, Mar 1991.
- [124] Yi Sun, Ole Brauckmann, Donald R. Nixdorf, Arno Kentgens, Michael Garwood, Djaudat Idiyatullin, Arend Heerschap. Imaging human teeth by phosphorus magnetic resonance with nuclear overhauser enhancement. *Scientific Reports*, 6(1), aug 2016.
- [125] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke. Inception-v4, inception-resnet and the impact of residual connections on learning. *CoRR*, abs/1602.07261, 2016.
- [126] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott E. Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich. Going deeper with convolutions. *CoRR*, abs/1409.4842, 2014.

- [127] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, Zbigniew Wojna. Rethinking the inception architecture for computer vision. *CoRR*, abs/1512.00567, 2015.
- [128] K. Tahara, Z. W. Luo, T. Odashima, M. Onishi, S. Hosoe, A. Kato. Dynamic control and simulation of human musculo-skeletal model. *SICE 2004 Annual Conference*, wolumen 3, strony 2395–2398 vol. 3, Aug 2004.
- [129] Nima Tajbakhsh, Jae Y. Shin, Suryakanth R. Gurudu, R. Todd Hurst, Christopher B. Kendall, Michael B. Gotway, Jianming Liang. Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE Transactions on Medical Imaging*, 35(5):1299–1312, may 2016.
- [130] Gavin Taylor, Ryan Burmeister, Zheng Xu, Bharat Singh, Ankit Patel, Tom Goldstein. Training neural networks without gradients: A scalable ADMM approach. *CoRR*, abs/1605.02026, 2016.
- [131] Joshua B. Tenenbaum, Vin de Silva, John C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319, 2000.
- [132] Laurens van der Maaten, Geoffrey Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9:2579–2605, 2008.
- [133] H. T. van Schie, E. M. Bakker. Structure-related echoes in ultrasonographic images of equine superficial digital flexor tendons. *Am. J. Vet. Res.*, 61(2):202–209, Feb 2000.
- [134] H. T. van Schie, E. M. Bakker, A. M. Jonker, P. R. van Weeren. Computerized ultrasonographic tissue characterization of equine superficial digital flexor tendons by means of stability quantification of echo patterns in contiguous transverse ultrasonographic images. *Am. J. Vet. Res.*, 64(3):366–375, Mar 2003.
- [135] H. T. M. van Schie, R. J. de Vos, S. de Jonge, E. M. Bakker, M. P. Heijboer, J. A. N. Verhaar, J. L. Tol, H. Weinans. Ultrasonographic tissue characterisation of human achilles tendons: quantification of tendon structure through a novel non-invasive approach. *British Journal of Sports Medicine*, 44(16):1153–1159, aug 2009.

- [136] Juha Vesanto, Johan Himberg, Esa Alhoniemi, Juha Parhankangas. Self-organizing map in matlab: the som toolbox. In *Proceedings of the Matlab DSP Conference*, strony 35–40, 2000.
- [137] D. Wang, A. Khosla, R. Gargya, H. Irshad, A. H. Beck. Deep Learning for Identifying Metastatic Breast Cancer. *ArXiv e-prints*, jun 2016.
- [138] Michael B. W. Wolfe, Susan R. Goldman. Use of latent semantic analysis for predicting psychological phenomena: Two issues and proposed solutions. *Behavior Research Methods, Instruments, & Computers*, 35(1):22–31, Feb 2003.
- [139] David H. Wolpert. Stacked generalization. *Neural Networks*, 5:241–259, 1992.
- [140] Robert H. Wurtz. Recounting the impact of hubel and wiesel. *The Journal of Physiology*, 587(12):2817–2823, jun 2009.
- [141] Saining Xie, Ross B. Girshick, Piotr Dollár, Zhuowen Tu, Kaiming He. Aggregated residual transformations for deep neural networks. *CoRR*, abs/1611.05431, 2016.
- [142] Guang Yang, Benjamin B. Rothrauff, Rocky S. Tuan. Tendon and ligament regeneration and repair: Clinical relevance and developmental paradigm. *Birth Defects Research Part C: Embryo Today: Reviews*, 99(3):203–222, sep 2013.
- [143] Matthew D. Zeiler, Rob Fergus. Visualizing and understanding convolutional networks. *CoRR*, abs/1311.2901, 2013.
- [144] Wei Zhang, Kunio Doi, Maryellen L. Giger, Yuzheng Wu, Robert M. Nishikawa, Robert A. Schmidt. Computerized detection of clustered microcalcifications in digital mammograms using a shift-invariant artificial neural network. *Medical Physics*, 21(4):517–524, apr 1994.

Dodatek A

AchillesDL: System komputerowego wspomagania oceny gojenia ścięgien i więzadeł