

WOJSKOWA AKADEMIA TECHNICZNA

im. Jarosława Dąbrowskiego



ROZPRAWA DOKTORSKA

**Strategie przygotowania obrazów pisma odręcznego na potrzeby
rozpoznawania choroby Parkinsona z wykorzystaniem sieci konwolucyjnej**

mgr inż. Kamila Bialek

Promotor: dr hab. inż. Jacek Jakubowski

Warszawa 2024

„Bez serc, bez ducha, — to szkieletów ludy!
Młodości! dodaj mi skrzydła!
Niech nad martwym wzlecę światem
W rajską dziedzinę uludy:
Kędy zapal tworzy cudu,
Nowości potrząsa kwiatem
I obleka w nadziei złote malowidła!”
~Adam Mickiewicz, Oda do Młodości

Dziękuję promotorowi dr hab. inż. Jackowi Jakubowskiemu za nieocenione wsparcie, pomoc merytoryczną i wszystkie cenne rady, które przyczyniły się do powstania niniejszej rozprawy. Dziękuję za poświęcony czas, cierpliwość i wiarę w powstanie tej pracy.

Dziękuję Pani prof. dr hab. n. med. Annie Kosterze-Pruszczyk za możliwość przeprowadzenia badań klinicznych z udziałem pacjentów Katedry i Kliniki Neurologii WUM oraz Pani dr hab. n. med. Annie Potulskiej-Chromik za cenną pomoc i nadzór medyczny podczas realizacji badań.

Dziękuję mojemu mężowi Rafałowi za ogromną motywację, wyrozumiałość i wsparcie oraz synowi Władowi, za wspólny czas spędzony z tatą, który dał mi przestrzeń i możliwość skoncentrowania się na pracy.

Dziękuję rodzicom za bezgraniczną wiarę w moje umiejętności, wychowanie mnie w świetle wartości edukacji i pokazywanie mi możliwości płynących z wykształcenia.

Streszczenie

Rozprawa doktorska dotyczy zastosowania nowoczesnych metod uczenia maszynowego w rozpoznawaniu choroby Parkinsona na podstawie obrazów pisma odręcznego uzyskanych z wykorzystaniem tabletu graficznego zastosowanego jako urządzenie pomiarowe. Celem pracy było zbadanie zagadnienia wpływu procesu przygotowania obrazów pisma na dokładność systemu rozpoznawania choroby Parkinsona wykorzystującego jako klasyfikator konwolucyjną sieć neuronową. Proces ten obejmował zaplanowanie eksperymentów rejestracji próbek pisma, wybór sposobu konwersji zarejestrowanych danych do postaci obrazów oraz przetworzenie uzyskanych obrazów na potrzeby wykorzystania sieci konwolucyjnej jako klasyfikatora.

Najistotniejszym aspektem uzyskanych wyników było stwierdzenie wpływu wydłużonego czasu pisania na manifestowanie zmian w piśmie wynikających z choroby. W przeciwieństwie do wcześniejszych badań obecnych w literaturze, które opierały się głównie na pojedynczych literach, słowach lub rysunkach, niniejsza rozprawa uwzględniała próbki składające się z kilku pełnych zdań zapisywanych przez badanego w ramach jednej sesji rejestracyjnej. Takie podejście pozwoliło nie tylko na uwypuklenie charakterystycznych zmian grafomotorycznych wynikających z choroby, ale także na zwiększenie liczby danych treningowych i tym samym poprawę zdolności uogólniania modelu. Dodatkowo zastosowano technikę wykorzystującą różne wzorce pisma jako dane uczące dla jednej sieci, co stanowi innowacyjne rozwiązanie w analizie pisma w kontekście rozpoznawania choroby Parkinsona. Dzięki temu model mógł skupić się na kluczowych zmianach wynikających z choroby, niezależnie od konkretnej treści zapisywanych zdań.

W niniejszej pracy zaprezentowano wyniki trzech podstawowych eksperymentów z wykorzystaniem sieci AlexNet do klasyfikacji osób chorych i zdrowych. Pierwszy eksperyment opierał się na analizie pojedynczego zdania, zgodnie z podejściem obecnym w literaturze. Następnie sprawdzono skuteczność klasyfikacji na podstawie wszystkich zdań zapisanych przez badaną osobę, co znacząco poprawiło wyniki. W ostatnim eksperymencie analizowano pojedyncze słowa, co dodatkowo zwiększyło dokładność rozpoznania choroby. Dodatkowo badano wpływ dostosowania głębokości uczenia oraz zastosowania bardziej złożonych sieci konwolucyjnych.

Uzyskane wyniki potwierdziły skuteczność zaproponowanego podejścia i jego potencjał w usprawnieniu wczesnej diagnozy choroby Parkinsona z wykorzystaniem technologii sztucznej inteligencji.

Słowa kluczowe: konwolucyjne sieci neuronowe, klasyfikacja, Choroba Parkinsona, diagnostyka, pismo odręczne

Abstract

This doctoral dissertation focuses on the application of modern machine learning methods in recognizing Parkinson's disease based on handwritten images obtained using a graphic tablet as a measurement device. The aim of the study was to investigate the impact of the handwriting image preparation process on the accuracy of a Parkinson's disease recognition system utilizing a convolutional neural network as a classifier. This process involved planning handwriting sample registration experiments, selecting a method for converting recorded data into images, and processing the obtained images for use in a convolutional network-based classifier.

The most significant aspect of the obtained results was the determination of the impact of prolonged writing time on the manifestation of handwriting changes resulting from the disease. Unlike previous studies in the literature, which mainly focused on individual letters, words, or drawings, this dissertation considered samples consisting of multiple full sentences written by a participant in a single registration session. This approach not only highlighted characteristic graphomotor changes caused by the disease but also increased the amount of training data, thereby improving the model's generalization ability. Additionally, a technique utilizing different handwriting patterns as training data for a single network was applied, representing an innovative approach to handwriting analysis in the context of Parkinson's disease recognition. This method enabled the model to focus on key disease-related handwriting changes regardless of the specific content of the written sentences.

This study presents the results of three key experiments using the AlexNet network to classify individuals as either Parkinson's patients or healthy controls. The first experiment analyzed a single sentence, following the approach commonly used in the literature. The second experiment evaluated classification performance based on all sentences written by each participant, significantly improving accuracy. The final experiment analyzed individual words, further enhancing the disease recognition performance. Additional studies investigated the effect of adjusting the network's learning depth and employing more complex convolutional neural networks.

The obtained results confirmed the effectiveness of the proposed approach and its potential for improving early diagnosis of Parkinson's disease using artificial intelligence technology.

Keywords: convolutional neural networks, classification, Parkinson's disease, diagnostics, handwriting

Spis treści

1. Wstęp.....	9
1.1. Charakterystyka problematyki wsparcia diagnostyki choroby Parkinsona	9
1.2. Cel i tezy rozprawy	13
1.3. Struktura rozprawy.....	15
2. Rozpoznawanie choroby Parkinsona na podstawie próbek pisma – stan zagadnienia.....	17
2.1. Parametryzacja pisma	17
2.2. Stan pacjentów	19
2.3. Wzorce pisma.....	20
2.4. Konwolucyjne sieci neuronowe – aktualny kierunek badań.....	22
3. Konwolucyjne sieci neuronowe – charakterystyka narzędzia	29
3.1. Operacja splotu	30
3.2. Nieliniowa funkcja aktywacji ReLu i pooling	32
3.3. Warstwy w pełni połączone - Fully Connected Layers	34
3.4. Uczenie sieci CNN.....	34
3.5. Augmentacja danych.....	36
3.6. Transfer learning.....	37
4. Rejestracja pisma odręcznego.....	39
4.1. Charakterystyka materiału badawczego	39
4.2. Tablet graficzny jako narzędzie do rejestracji pisma odręcznego	40
5. Wstępne przetwarzanie danych	43
5.1. Podział całych rejestracji na poszczególne zdania.....	43
5.2. Podział zdań na słowa.....	44
5.3. Generacja obrazów na podstawie zarejestrowanych danych	45
6. Organizacja badań eksperymentalnych	48
6.1. Walidacja krzyżowa.....	48
6.2. Miary jakości klasyfikacji.....	49
7. Rozpoznawanie choroby Parkinsona na podstawie pojedynczego zdania	53
7.1. Metodyka badań.....	53
7.1.1. Struktura sieci AlexNet	53
7.1.2. Przygotowanie danych	54
7.1.3. Ekstrakcja cech.....	55
7.1.4. Dostrajanie	56
7.1.5. Uczenie sieci.....	57

7.2. Wyniki eksperymentu.....	57
7.2.1. Rozpoznawanie na podstawie pierwszego zdania	57
7.2.2. Rozpoznawanie na podstawie poszczególnych zdań.....	59
8. Rozpoznawanie choroby Parkinsona na podstawie sekwencji zdań.....	63
8.1. Metodyka badań	63
8.1.1. Strategia decyzyjna	63
8.1.2. Augmentacja pisma odręcznego	64
8.1.3. Uczenie sieci	65
8.2. Wyniki eksperymentu.....	65
8.2.1. Rozpoznawanie na podstawie zarejestrowanych zdań	65
8.2.2. Rozpoznawanie z wykorzystaniem augmentacji danych.....	68
9. Rozpoznawanie choroby Parkinsona na podstawie słów z sekwencji zdań	70
9.1. Metodyka badań	70
9.1.1. Przygotowanie danych.....	70
9.1.2. Uczenie sieci	72
9.2. Wyniki eksperymentu.....	73
9.2.1. Rozpoznawanie na podstawie sekwencji słów	73
9.2.2. Wpływ prezentacji danych na obrazie na wynik klasyfikacji – porównanie wyników	76
10. Prace optymalizacyjne w zakresie doboru klasyfikatora	79
10.1. Dobór głębokości dostrojenia sieci AlexNet.....	79
10.2. Wykorzystanie innych sieci konwucyjnych	83
11. Podsumowanie i wnioski końcowe	86
Bibliografia	89
ZAŁĄCZNIK 1 Szczegółowe informacje kliniczne i demograficzne na temat uczestników badania	94
ZAŁĄCZNIK 2 Autorskie oprogramowanie do rejestracji pisma.....	96

1. Wstęp

1.1. Charakterystyka problematyki wsparcia diagnostyki choroby Parkinsona

Choroba Parkinsona jest postępującym zaburzeniem neurodegeneracyjnym charakteryzującym się osłabieniem funkcji motorycznych. Obecnie jest to druga po chorobie Alzheimera najczęstsza choroba neurozwyrodnieniowa [1]. Jednocześnie uważana jest za najszybciej rozwijającą się chorobę neurologiczną na świecie [2]. Szacuje się, że cierpi na nią około 0,1 % całej populacji. Wśród osób powyżej 60. roku życia odsetek ten wynosi około 1 %, i wzrasta nawet do 3 % wśród osób powyżej 80. roku życia [3]. Prognozuje się, że do 2030 roku liczba osób z chorobą Parkinsona w Europie Zachodniej osiągnie ponad 9 milionów [4]. Mimo, że dotyczy ona głównie osoby starsze (średni wiek pojawienia się pierwszych objawów wynosi 58 lat), a częstość jej występowania rośnie wraz z wiekiem, to zdarzają się przypadki zachorowań poniżej 50 roku życia [5].

Głównymi objawami choroby Parkinsona są: drżenie spoczynkowe, spowolnienie ruchowe (bradykinezja), sztywność kończyn oraz niestabilność postawy ciała [6]. Mimo nieokreślonej etiologii, mechanizmy będące podłożem występujących dolegliwości i objawów tej patologii są dobrze znane. Zasadniczym zjawiskiem jest obumieranie komórek istoty czarnej, które prowadzi do spadku poziomu dopaminy w prążkowie. Dopamina pełni funkcję chemicznego neuroprzekaźnika wspomagającego system nerwowy w procesie kontroli aktywacji mięśni [7]. Dlatego współczesna diagnostyka choroby Parkinsona bazuje na bezpośrednim obrazie klinicznym i wywiadzie chorobowym, których celem jest stwierdzenie współistnienia wymienionych objawów, a następnie na wprowadzeniu leczenia farmakologicznego i obserwacji jego wpływu na pacjenta. Zadowolająca odpowiedź organizmu na terapię dopaminergiczną, potwierdza rozpoznanie choroby. Jednakże wymienione objawy występują dopiero w późniejszym stadium choroby, kiedy jej rozpoznanie jest stosunkowo proste. Do wczesnych objawów można natomiast zaliczyć: zaparcia, problemy ze snem, hiposmię (zmniejszoną zdolność do wyczuwania zapachów), niewyraźny ból barku czy depresję [6, 8]. Ze względu na fakt, że objawy te nie są niepokojące i mogą dotyczyć wielu różnych chorób, rozpoznanie wczesnych stadiów choroby Parkinsona może być trudne, o czym świadczy duże opóźnienie (średnio 10 lat) między pierwszym zauważalnym objawem a momentem postawienia diagnozy [9]. Opóźnienia w diagnozie są szczególnie częste, gdy nie występuje drżenie lub gdy choroba dotyczy głównie nóg. Jednocześnie błędne rozpoznanie prowadzi do

tego, że pacjent jest leczony bez uzasadnienia lekami mającymi silne działania niepożądane. Stąd dużą wagę przywiązuje się do diagnostyki wykorzystującej zestaw zwalidowanych testów klinicznych oceniających pacjenta z wykorzystaniem tzw. skali UPDRS [10]. Poszczególne komponenty oceny choroby Parkinsona według skali UPDRS przedstawiono na rysunku 1.1.

UPDRS – Unified Parkinson's Disease Rating Scale - obejmuje 4 komponenty o łącznej liczbie 50 sekcji, punktowanych od 0 do 4, gdzie 0 oznacza brak upośledzenia a 4 upośledzenie ciężkie, - wynik całkowity jest sumą wyników z poszczególnych sekcji; 0 oznacza w 100 % przypadek zdrowy	
Komponent I Pozaruchowe aspekty umiejętności w zakresie czynności codziennych - ocena zaburzeń funkcji poznawczych, stanu psychiki, snu, zmęczenia, czynności fizjologicznych	Komponent II Ruchowe aspekty codziennego funkcjonowania - ocena problemów z komunikacją, jedzeniem, ubieraniem się, chodzeniem, równowagą i pisanie
Komponent III Objawy ruchowe - ocena jakości mowy, wyrazy twarzy, motoryki kończyn i palców, drżenia zamierzonego i spoczynkowego, postawy, wstawania z krzesła, chodu	Komponent IV Powikłania ruchowe - ocena stopnia zaawansowania i czasu trwania: dyskinez, dystonii, fluktuacji stanu sprawności ruchowej

Rys.1.1. Komponenty oceny choroby Parkinsona w badaniach klinicznych

Opublikowana po raz pierwszy w 1987 roku ujednolicona skala UPDRS (ang. Unified Parkinson's Disease Rating Scale) jest skalą oceniającą 4 komponenty rozwoju choroby Parkinsona w gradacji pięciostopniowej, w której wyższe zakresy odpowiadają bardziej zaawansowanym stadiom [10]. Na podstawie wywiadu oraz wykonanych testów określa się z jej wykorzystaniem stan intelektualny i zaburzenia nastroju (komponent I), jakość codziennego życia (komponent II), funkcje motoryczne (komponent III) i powikłania leczenia (komponent IV). Dzięki przeprowadzonemu procesowi walidacji, skala UPDRS jest formalnym narzędziem stosowanym na całym świecie jako podstawowy wskaźnik w diagnostyce, monitorowaniu progresji i w badaniach nowych metod leczenia neuroprotektynowego.

W zakresie oceny objawów ruchowych jest to jednak narzędzie mało obiektywne i wymaga znacznego doświadczenia osób przeprowadzających testy. Na przykład jednym z testów oceniających funkcje motoryczne jest test szybkiego stukania palcami RFT (ang. *rapid finger tapping*). Polega on na jak najszybszym wykonywaniu co najmniej 10 stuknięć palcami z użyciem tylko palca wskazującego i kciuka przy jednoczesnym jak największym zakresie ruchu (amplitudzie). Ocenie wyłącznie wizualnej, w skali od 0 do 4 podlega jakość wykonywania zadania przez każdą z rąk, opisywana liczbą zaburzeń regularności ruchu, tempem spowolnienia i zmniejszaniem się amplitudy w czasie. Podobnie niskim poziomem obiektywności charakteryzuje się ta część testu klinicznego, w której w ramach konwersacji ocenia się mowę pacjenta. Zmiany w głosie, osób z chorobą Parkinsona wynikają z jej wpływu na proces oddychania, fonacji i artykulacji. Ocenie klinicznej, również według skali od 0 do 4 podlegają: siła głosu, brzmienie (prozodia), zrozumiałość, dykcja, powtarzanie się sylab

(palialia) i zlewania się sylab (tachyfemia). Pozwala to jedynie na subiektywne wypracowanie oceny stopnia upośledzenia mowy badanego przez osobę przeprowadzającą test. Kolejnym przykładem braku obiektywności badań jest ocena wyrazu twarzy. Z powodu pogarszającej się szybkości działania i koordynacji mięśni twarzy, wśród osób dotkniętych zaburzeniami neurodegeneracyjnymi obserwuje się trudności z wyrażaniem emocji na twarzy, która przyjmuje wyraz amimiczny (brak mimowolnych, spontanicznych ruchów mimicznych, maskowatość twarzy lub tzw. twarz „pokerowa”). Instrukcja badania klinicznego wyrazu twarzy wg. UPDRS przewiduje jedynie obserwację twarzy pacjenta w ciszy oraz w trakcie mówienia, której celem jest określenie stopnia zubożenia mimiki na podstawie oceny częstotliwości mrugania oraz występowania niedomykania się ust w trakcie spoczynku. Charakterystycznym symptomem choroby Parkinsona, chociaż występującym jedynie w 75 % przypadków, jest drżenie kończyn [11]. W praktyce klinicznej, celem wypracowania wskaźnika liczbowego oceny w pięciostopniowej skali UPDRS, drżenie ocenia się jedynie na podstawie wizualnej obserwacji amplitudy ruchu względem wartości progowych 0 cm, 1 cm, 3 cm i 10 cm. Brak pełnej obiektywności oceny objawów w przeprowadzonych badaniach klinicznych powoduje, że chorobę Parkinsona diagnozuje się nieprawidłowo nawet w 15 % przypadków [12]. Z tego względu współczesna neurologia zgłasza zapotrzebowanie na narzędzia umożliwiające pozyskiwanie informacji o stanie pacjenta, archiwizację tej informacji celem określenia stopnia zaawansowania schorzenia i badania wpływu środków farmakologicznych oraz na narzędzia wspomagające diagnozę. Dlatego szereg ośrodków naukowych podejmuje współpracę ze środowiskiem medycznym celem wsparcia istniejącego procesu diagnozowania i oceny progresji choroby obiektywnymi metodami inżynierskimi, poszukując wskaźników będących biomarkerami choroby, zwłaszcza na wczesnym etapie jej rozwoju.

W literaturze prezentowanych jest wiele udokumentowanych wyników badań prowadzonych w obszarze analizy różnych materiałów biometrycznych, które można uzyskać w trakcie standardowych badań klinicznych wykorzystując znane urządzenia do ich obiektywnej rejestracji. Historycznie pierwszymi podejściami były urządzenia wykorzystujące sensory ruchu, przede wszystkim akcelerometry do pomiarów chodu i drżenia kończyn [13, 14]. Akcelerometryczna metoda zbierania danych dotyczących drżenia, pomimo swej prostoty, jest obecnie uważana za najlepszą ze względu na wysoką czułość powszechnie dostępnych sensorów, głównie sensorów MEMS [15]. Obserwowany w ostatnich dziesięcioleciach rozwój metod rozpoznawania mowy i mówców sprawił, że szeroką kategorię prac badawczych poświęconych diagnostyce w chorobie Parkinsona stanowią prace prowadzone w zakresie analizy mikrofonowych rejestracji głosu [16-18]. Dzięki upowszechnieniu metod widzenia maszynowego, w tym metod opartych na głębokim uczeniu, zauważalną grupę prac stanowią również prace prezentujące wykorzystanie technik wizyjnych, zarówno do oceny maskowatości twarzy, jak i funkcji motorycznych [19-22].

Materiałem biometrycznym, który niesie ze sobą duży potencjał diagnostyczny, a jednocześnie jego rejestracja, przy zastosowaniu odpowiedniego urządzenia, pozbawiona jest wielu problemów występujących przy rejestracji innych rodzajów danych, jest pismo odręczne. Pisanie, czyli tworzenie pisma ręcznego jest czynnością prostą i naturalną. Jego rejestracja nie wymaga specjalnych warunków środowiskowych, wycucia czasu ani wielu wyczerpujących powtórzeń. Ponieważ choroba Parkinsona jest zaburzeniem neurologicznym wywołującym problemy ze sprawnością aparatu ruchu, pojawienie się któregośkolwiek z motorycznych objawów wpływa na grafomotorykę, czyli umiejętność pisania odręcznego. Proces ten wymaga zdolności manualnej i prawidłowej percepcji wzrokowej, gdyż polega na precyzyjnych, sekwencyjnych ruchach ręką z jednoczesną koordynacją wzrokowo-ruchową [23]. Przykładowo, spowolnienie ruchowe, może sprawiać, że osoba chora będzie potrzebowała dłuższego czasu niż osoba zdrowa na wykonanie tego samego zadania piśmienniczego. Drżenie rąk, czyli ich mimowolne, oscylacyjne ruchy mogą spowodować, że pismo będzie nieregularne, niewyraźne lub poszarpane. Dodatkowo, jedną z najbardziej charakterystycznych cech pisma osób z chorobą Parkinsona jest zjawisko mikrografii polegające na stopniowym zmniejszaniu rozmiaru czcionki w trakcie pisanie [24]. Prezentowane w literaturze badania potwierdzają, że osoby chore piszą wolniej, niż osoby zdrowe z grupy kontrolnej [25, 26]. Ich pismo charakteryzuje się większą fluktuacją linii bazowej [25], jest bardziej chaotyczne i nieregularne [26]. Dodatkowo osoby chore nie kontrolują siły nacisku podczas pisanie w takim stopniu jak osoby zdrowe [27]. W związku z powyższym zmiany w piśmie oraz problemy z pisaniem mogą być pierwszymi zauważalnymi oznakami występowania choroby Parkinsona i jednocześnie stać się podstawą do automatyzacji procesu rozpoznawania choroby na wczesnym etapie jej rozwoju z wykorzystaniem technicznych narzędzi do rejestracji pisma.

Analizując pismo pod kątem możliwości stwierdzenia na jego podstawie choroby Parkinsona należy jednak mieć na uwadze fakt, że sposób pisanie jest bardzo osobniczy i zależy od wielu innych czynników, nie koniecznie związanych z chorobą jak np. od zdolności plastycznych, prawo- lub leworęczności, sposobu trzymania długopisu, wieku czy po prostu poczucia estetyki osoby piszącej [23]. Dodatkowo, występowanie zmian w piśmie oraz ich nasilenie spowodowane chorobą będzie zależało od jej rodzaju i stadium. W związku z powyższym jednoznaczne wskazanie cech pisma, które mogłyby stać się nowymi biomarkerami choroby Parkinsona wydaje się być zadaniem bardzo trudnym. Tym niemniej, podkreślane w literaturze, obiecujące znaczenie analizy pisma odręcznego oraz obserwowany w ostatniej dekadzie rozwój metod uczenia maszynowego, w tym uczenia głębokiego do diagnozowania różnych chorób na podstawie przetwarzania obrazów medycznych – guzów mózgu, nowotworów piersi i skóry [28-32], stały się dla autora bodźcem do przeprowadzenia własnych badań i napisania niniejszej rozprawy, w której przedmiotem zainteresowania jest właśnie wykorzystanie pisma odręcznego w diagnostyce choroby Parkinsona.

Rozprawa powstała w efekcie prac realizowanych na Wydziale Elektroniki WAT w ramach zadania *Innowacyjne metody i algorytmy diagnostyki medycznej* na potrzeby projektu badań

podstawowych pn. *Badania podstawowe w obszarze technologii sensorowej z wykorzystaniem innowacyjnych metod przetwarzania*, finansowanych przez Ministerstwo Obrony Narodowej w ramach grantu GBMON/B9/13-996/2018. Kwalifikacją pacjentów do udziału w badaniach, zgodnie z wymaganiami właściwej komisji bioetycznej, zajął się zespół medyczny z Katedry i Kliniki Neurologii Warszawskiego Uniwersytetu Medycznego, współpracujący na mocy podpisanego porozumienia. W ramach współpracy opracowano stanowisko pomiarowe oraz procedurę zbierania danych odwołując się w swej istocie do specyfiki testów klinicznych zalecanych przez Międzynarodowe Towarzystwo Choroby Parkinsona i Zaburzeń Ruchu. Zapewniły one możliwość obiektywnego przetwarzania danych opisujących objawy ruchowe oraz próbek głosu i pisma odręcznego ocenianych w ramach III i II komponentu stosowanej w praktyce klinicznej skali UPDRS.

1.2. Cel i tezy rozprawy

Niniejsza rozprawa doktorska skupia się na zagadnieniach zastosowania współczesnych metod uczenia maszynowego w rozpoznawaniu choroby Parkinsona na podstawie obrazów pisma odręcznego uzyskanych z wykorzystaniem tabletu graficznego. **Celem pracy było zbadanie zagadnienia wpływu procesu przygotowania obrazów pisma na dokładność systemu rozpoznawania choroby Parkinsona wykorzystującego jako klasyfikator konwolucyjną sieć neuronową.** Przez proces przygotowania obrazów rozumiano ogół czynności obejmujących zaplanowanie eksperymentów rejestracji próbek pisma, wybór sposobu konwersji zarejestrowanych danych do postaci obrazów oraz przetworzenie uzyskanych obrazów na potrzeby wykorzystania sieci konwolucyjnej jako klasyfikatora. Proponowane rozwiązania mają na celu wspomóc proces wczesnej identyfikacji choroby Parkinsona z wykorzystaniem technologii sztucznej inteligencji.

Znaczna część badań obecnych w literaturze podjętych przez innych autorów w zakresie rozpoznawania choroby Parkinsona z wykorzystaniem grafik tworzonych przez pacjentów, bazuje na powstałych kilka lat wcześniej i udostępnionych zespołom badawczym bazach danych. Większość z nich zawiera wyłącznie obrazy odręcznie tworzonych rysunków np. tzw. spiral Archimedes'a [33-35], meandrów [35] czy trójkątnych fal [36]. Natomiast dostępne bazy danych zawierające próbki pisma są stosunkowo nieliczne i dodatkowo oferują jedynie zapisy: pojedynczych liter, pojedynczych słów lub jednego pojedynczego zdania. W pracach wykorzystujących powyższe wzorce autorzy skupiali się zatem wyłącznie na analizie danych zarejestrowanych w krótkim czasie. We współczesnej literaturze brakuje prac, które uwzględniałyby zmiany w piśmie zachodzące w wyniku przedłużonego procesu pisania z powodu spodziewanego szybszego zmęczenia osób chorych podjętym wysiłkiem grafomotorycznym. Z tego względu, już na etapie opracowywania koncepcji badań prezentowanych w niniejszej rozprawie, w ramach planowania eksperymentu zaproponowano podejście polegające na pobraniu od jednej osoby badanej więcej niż jednego zdania w trakcie tej samej sesji rejestracyjnej. Prezentowane wyniki dotyczą badań przeprowadzonych

z wykorzystaniem próbek pisma zawierających pięć zdań o jednakowej treści zapisanych jedno pod drugim. Tak przygotowane dane pozwalają uwzględnić w procesie uczenia sieci nie tylko modyfikacje pisma wynikające z obecności choroby, ale również uzyskać zwielokrotnienie liczby danych uczących osiągane zazwyczaj technikami augmentacyjnymi. Dzięki możliwości wykorzystania każdego zdania lub słowa oddzielnie jako odrębnej próbki uczącej, powstały tym sposobem również warunki do poprawy zdolności uogólniania. Użycie różnych wzorców pisma (różnych wyrazów), jako danych uczących jedną sieć, jest techniką niewykorzystywaną w obecnych badaniach nad analizą pisma. Natomiast pozwala ono na uniezależnienie się procesu uczenia od wzorca pisma i jednocześnie skupienie się na istotnych zmianach wynikających z choroby bez względu na zapisywaną treść.

Aby zrealizować cel pracy rozwiązano następujące zadania badawcze:

1. Opracowano elementy stanowiska badawczego w części przeznaczonyj do rejestracji próbek pisma odręcznego pacjentów za pomocą tabletu graficznego.
2. Zaplanowano i przeprowadzono eksperymenty rejestracyjne próbek pisma z udziałem grupy badawczej pacjentów ze zdiagnozowaną chorobą Parkinsona oraz grupy kontrolnej osób zdrowych.
3. Zaproponowano różne strategie przygotowania danych obrazowych dla sieci konwolucyjnej uwzględniające możliwości wykorzystania próbek pisma o przedłużonym czasie rejestracji.
4. Zaproponowano metodykę podejmowania decyzji przez sieć konwolucyjną dla przypadków wykorzystania jako dane wejściowe sekwencji obrazów zdań i sekwencji obrazów słów.
5. Zaplanowano i przeprowadzono eksperymenty numeryczne wykorzystujące walidację krzyżową celem uzyskania wiarygodnej oceny procesu klasyfikacji dla zaproponowanych strategii.

W efekcie analizy teoretycznej oraz uzyskanych wyników prac badawczych prowadzonych celem uzyskania dużej dokładności rozpoznawania choroby Parkinsona na podstawie próbek pisma z wykorzystaniem sieci konwolucyjnych, uważa się za uzasadnione następujące tezy rozprawy:

- 1) **użycie sekwencji obrazów zdań daje lepsze wyniki klasyfikacji binarnej w diagnostyce choroby Parkinsona z wykorzystaniem konwolucyjnej sieci neuronowej niż użycie obrazu zdania pojedynczego,**
- 2) **możliwa jest dalsza poprawa dokładności klasyfikacji binarnej w diagnostyce choroby Parkinsona z wykorzystaniem konwolucyjnej sieci neuronowej na drodze podziału obrazów sekwencji zdań na obrazy sekwencji słów.**

1.3. Struktura rozprawy

Z uwagi na charakter postawionych tez, w pracy zasadniczo założono wykorzystanie jednej sieci neuronowej, która pełniła rolę narzędzia do ich weryfikacji w ramach zrealizowanych eksperymentów numerycznych. Wykorzystano do tego celu ogólnie dostępną i szeroko stosowaną sieć AlexNet, przy czym zrealizowano szereg badań uzupełniających związanych ze zmianą głębokości dostrojenia, których celem było pełne wykorzystanie jej możliwości. Jako dodatkowe, do celów porównawczych, zostały również wykonane i zaprezentowane wyniki eksperymentów z wykorzystaniem innych, bardziej rozbudowanych modeli jak VGG16, GoogleNet i ResNet18, przy czym w ich przypadku proces dostrojenia dotyczył tylko ostatniej warstwy pełnego połączenia.

Rozprawa podzielona jest na 10 rozdziałów. Rozdział pierwszy stanowi wprowadzenie w tematykę rozprawy. Przedstawiono w nim cel oraz tezy niniejszej dysertacji.

W rozdziale drugim przeprowadzono analizę literatury dotyczącej rozpoznawania choroby Parkinsona na podstawie pisma odręcznego. Analizę dotychczasowych osiągnięć w tym temacie przeprowadzono pod kątem: stosowanych narzędzi, stanu pacjentów oraz używanych wzorców pisma. Scharakteryzowano obecnie najczęściej wykorzystywaną w tego typu badaniach bazę danych PaHaW oraz omówiono wyniki najnowszych badań uzyskanych z wykorzystaniem konwolucyjnych sieci neuronowych na podstawie próbek pisma pochodzących z tej bazy.

Rozdział trzeci stanowi krótkie wprowadzenie teoretyczne do konwolucyjnych sieci neuronowych możliwych do wykorzystania w podejmowaniu decyzji diagnostycznych na podstawie obrazów pisma. Omówiono w nim podstawowe nazewnictwo, budowę sieci oraz zagadnienia związane z technikami wykorzystywania tego typu rozwiązania w zadaniach klasyfikacji. Rozdział ten stanowi swego rodzaju bazę wiedzy na temat własności tego narzędzia i metodyki jego wykorzystania w dalszej części pracy.

W rozdziale czwartym zawarto charakterystykę osób, od których pozyskano próbki pisma odręcznego w ramach współpracy z Kliniką i Katedrą Neurologii Warszawskiego Uniwersytetu Medycznego. W tym rozdziale zaprezentowano stanowisko do rejestracji pisma wraz z specyfikacją tabletu graficznego i opisem przebiegu badania.

W rozdziale piątym omówiono metodykę generacji obrazów na podstawie danych pozyskanych w wyniku rejestracji pisma za pomocą tabletu graficznego.

Rozdział szósty stanowi wprowadzenie do części eksperymentalnej. Zawarto tutaj definicje wykorzystanych w badaniach miar jakości klasyfikacji, które posłużyły do oceny procesu rozpoznawania choroby.

Rozdział siódmy, ósmy i dziewiąty stanowią właściwą część eksperymentalną. Zamieszczono tutaj wyniki badań dążących do potwierdzenia postawionych na początku niniejszej rozprawy tez. Każdy z rozdziałów został podzielony na dwie części: metodykę badań, gdzie omówiono sposoby i narzędzia wykorzystane do przeprowadzenia eksperymentu oraz uzyskane wyniki.

W rozdziale siódmym zawarto wyniki klasyfikacji binarnej osób chorych na chorobę Parkinsona i osób zdrowych na podstawie pisma odręcznego przedstawiającego pojedyncze zdanie. Eksperyment został przeprowadzony z wykorzystaniem różnych technik *transfer learningu* i miał na celu odniesienie się do wyników podobnych badań zamieszczonych w kluczowych publikacjach dotyczących podjętego tematu. Poruszono tutaj również kwestię wpływu czasu trwania wysiłku grafomotorycznego na wynik rozpoznawania choroby Parkinsona. Wyniki zaprezentowane w tym rozdziale posłużyły jako odniesienie dla wyników kolejnych badań.

W rozdziale ósmym zaprezentowano badanie, w którym jako dane uczące użyto obrazy prezentujące wyodrębnione zdania pochodzące z zapisu ich sekwencji. Omówiono tutaj strategię integracji decyzji podejmowanych na podstawie pojedynczych zdań oraz porównano wyniki z tymi uzyskanymi przy użyciu klasycznej augmentacji.

W rozdziale dziewiątym przedstawiona została klasyfikacja osób chorych na chorobę Parkinsona i osób zdrowych na podstawie próbek pisma przedstawiających pojedyncze słowa pochodzące z podziału zapisu pięciu zdań. Eksperyment przeprowadzono dla dwóch różnych metod prezentacji pisma w postaci obrazu, dla których przeanalizowano wpływ podziału zdań na słowa na wynik rozpoznawania choroby.

W rozdziale dziesiątym podjęto próbę poprawy wyniku klasyfikacji poprzez dostosowanie głębokości uczenia sieci AlexNet oraz wykorzystanie innych, bardziej rozbudowanych sieci konwolucyjnych.

Ostatni rozdział zawiera podsumowanie wyników całej rozprawy, wnioski płynące z przeprowadzonych badań, ich krytyczną ocenę oraz sugestie dotyczące kierunków dalszych badań w dziedzinie analizy pisma odręcznego z wykorzystaniem konwolucyjnych sieci neuronowych.

2. Rozpoznawanie choroby Parkinsona na podstawie próbek pisma – stan zagadnienia

Poniższa analiza literatury dotyczy badań nad pismem odręcznym pod kątem choroby Parkinsona. Nie odnoszono się tutaj do badań, w których przedmiotem były próbki odręcznie tworzonej grafiki np. okręgów czy spirali, które również mogą zawierać informację diagnostyczną. W literaturze można znaleźć wiele pozycji opisujących te zagadnienia. W pracy skupiono się wyłącznie na piśmie, które jako system graficznego wyrażania myśli, jest najbardziej uniwersalną i jednocześnie wysoko zindywidualizowaną cechą behawioralną człowieka [37]. Tworzenie pisma niosącego pewien logiczny komunikat jest również czynnością dużo bardziej naturalną niż rysowanie figur geometrycznych czy spiral. Ta uniwersalność pisma odręcznego spowodowała, że badania nad wpływem choroby Parkinsona na pismo odręczne przeprowadzano również w krajach, gdzie ludzie posługują się alfabetem innym niż łaciński. Znalaziono przykładowe prace dotyczące alfabetu: koreańskiego [38], cyrylicznego [39] i arabskiego [40].

2.1. Parametryzacja pisma

Jedną z najbardziej charakterystycznych zmian zauważalnych w piśmie osób chorych jest mikrografia, czyli zjawisko polegające na stopniowym zmniejszaniu wielkości pisma w trakcie pisania [24]. W pierwszych badaniach poświęconych analizie pisma osób cierpiących na chorobę Parkinsona, skupiano się właśnie na występowaniu tego zjawiska [41, 42]. W latach 90. zaczęto badać również inne, wynikające z obecności choroby, zmiany mogące pojawić się w piśmie odręcznym, [25, 43-46]. Początkowo autorzy prac wykorzystywali narzędzia analizy statystycznej, aby wskazywać różnice w piśmie i sposobie jego tworzenia u osób chorych i osób zdrowych [25, 26, 43, 44, 46-56]. W badaniach tych na podstawie zarejestrowanych, za pomocą urządzeń digitalizujących, próbek pisma obliczano niewielki zbiór parametrów, będących podstawą dalszej analizy. Najczęściej na podstawie otrzymanych danych przeprowadzano analizę wariancji (ANOVA) lub jej wielowymiarową odmianę (MANOVA). Metody te miały tu na celu wykazać różnice pomiędzy wartościami średnimi wektorów opisujących próbki pisma wykonywanego przez pacjentów z grupy osób chorych i zdrowych [26, 45, 48, 52-54]. Intensywny rozwój technik uczenia maszynowego spowodował, że w badaniach nad pismem odręcznym pod kątem wykrycia choroby Parkinsona zaczęto przeprowadzać automatyczną klasyfikację osób chorych i zdrowych za pomocą nowoczesnych algorytmów klasyfikacyjnych [57-68]. Prace te prezentują tzw. podejście klasyczne, inaczej nazywane inżynierią cech, które

polega na wskazaniu wektorów wartości liczbowych charakterystycznych dla populacji, bezpośrednio z wyników wstępnego przetwarzania danych eksperymentalnych, a następnie na zastosowaniu jednej z metod uczenia maszynowego do przeprowadzenia klasyfikacji. Do najczęściej wykorzystywanych klasyfikatorów można zaliczyć: SVM (ang. *Support Vector Machines*) [57, 64, 67, 68], KNN (ang. *K-Nearest Neighbors*) [59, 60, 62, 63], AdaBoost z estymatorem drzewa decyzyjnego [59-61] czy las losowy [60, 63]. Bez względu na użyte narzędzia analizy danych, powyższe badania wymagały przeprowadzenia parametryzacji pisma, czyli obliczenia zbioru parametrów, charakteryzujących pismo odręczne. Pojawienie się na rynku zaawansowanych urządzeń rejestrujących, takich jak tablety graficzne czy elektroniczne pióra [51], umożliwiło parametryzację nie tylko pisma, jako śladu widocznego na papierze, ale również procesu jego tworzenia, z uwzględnieniem jego kinematyki i dynamiki [69, 70]. Nowoczesne urządzenia tego typu rejestrują nie tylko trajektorię ruchu z odpowiednią częstotliwością próbkowania, ale również wartość siły nacisku w danej chwili czasu oraz kąt nachylenia pióra do powierzchni tabletu. Możliwość wykorzystania tej dodatkowej funkcjonalności tabletów graficznych w rejestracji próbek pisma zapoczątkowała nową erę w badaniach nad pismem odręcznym. Pojawiło się wiele publikacji dotyczących wyznaczania coraz większej liczby parametrów pisma, mających wskazywać na występowanie choroby Parkinsona. Zaczęto od określania rozmiaru czcionki na podstawie już nie tylko wyznaczanych wysokości [25, 43-50, 55, 56] czy długości kolejnych pociągnięć pióra [46, 48, 49, 55], ale również na podstawie długości trajektorii ruchu tych pociągnięć [55, 60, 61, 65]. Oprócz analizy cech statycznych przeprowadzano również analizę parametrów czasowych. Obliczany był ogólny czas pisania [25, 26, 43-50, 54, 58, 59, 62, 65, 66]. Ponadto pojawiło się kilka prac w których wyodrębniano czas, kiedy pióro dotyka powierzchni tabletu oraz czas, gdy pióro znajduje się nad powierzchnią tabletu [53, 57, 59, 65, 66, 68], a nawet stosunek obu tych czasów [56, 57, 59, 61, 65, 66, 68]. Takie podejście umotywowano tym, że czas trwania przerw między poszczególnymi pociągnięciami niesie dodatkową informację diagnostyczną, jako, że wiąże się on z czasem potrzebnym do zaplanowania następnego ruchu i rozmieszczenia przestrzennego, co może wskazywać na pogorszenie możliwości poznawczych. W [56] uwzględniony był dodatkowo czas reakcji, czyli odstęp czasu pomiędzy sygnałem startu a rozpoczęciem realizacji zadania piśmienniczego. Skupiano się również na pojedynczych parametrach dynamicznych jak: prędkość [26, 45, 52-54, 71], przyspieszenie [26, 43, 44, 47, 50, 71] oraz zryw [49, 71], definiowany jako miara zmiany przyspieszenia w czasie, będący według autorów dobrą miarą zniekształcenia ruchu. W wielu pracach uwzględniono więcej niż jeden z tych parametrów [55, 66, 68, 71]. Dodatkowo w [57-61, 64-66, 68] wyznaczono również parametry związane z prędkością i przyspieszeniem: NCV i NCA, zdefiniowane jako średnia liczba lokalnych ekstremów kolejno prędkości i przyspieszenia. W niektórych badaniach przeprowadzano również analizę nacisku [53, 55, 56, 59-61, 63, 66, 68]. Pojawiały się również inne ciekawe parametry pisma ręcznego oparte na entropii, energii sygnału [51, 58, 60, 61, 64] i rozkładzie sygnału na mody empiryczne [58, 67]. Uzyskanie pewnej liczby cech umożliwiło ilościową

ocenę nie tylko mikrografii [45, 55], ale również innych objawów choroby Parkinsona, jak spowolnienie ruchowe [26, 44, 54] czy drżenie rąk [45]. Jednak badania wskazują, że mikrografia występuje tylko u pewnej liczby pacjentów cierpiących na chorobę Parkinsona [24, 41, 55] a stopień nasilenia pozostałych objawów jest zależny od stadium choroby, poziomu trudności zadania piśmienniczego do wykonania podczas badania [47, 50], jednoczesnego wykonywania innych czynności [27, 72, 73] oraz od przyjmowanych leków [74, 75]. W związku z tym występowanie pojedynczych objawów w piśmie odręcznym czy nawet kilku jednocześnie nie umożliwia wiarygodnej klasyfikacji osób chorych. Dopiero wyznaczenie bardzo dużej liczby parametrów pisma i zastosowanie odpowiedniego algorytmu dyskryminującego umożliwiło znalezienie podzbioru cech odpowiedniego do przeprowadzenia zadowalającej klasyfikacji osób chorych [57-61]. Uzyskiwane tym sposobem dokładności rozpoznawania zależą od zastosowanych metod ekstrakcji i selekcji cech oraz metod klasyfikacji. Nowe możliwości przetwarzania pisma odręcznego dały współczesne rozwiązania głębokich sieci neuronowych. Na przykład rekurencyjne sieci neuronowe, specjalizujące się w przetwarzaniu danych sekwencyjnych, zastosowano w kilku pracach do analizy szeregów czasowych wyznaczanych parametrów [36, 76, 77]. Odmienne podejście, bazujące nie na wektorach cech, ale bezpośrednio na całych obrazach, w poszukiwaniu statycznej informacji wizualnej umożliwiają sieci konwolucyjne. Sieci te generują w sposób automatyczny (bez udziału człowieka) zbiór cech charakterystycznych dla analizowanych wzorców obrazowych, który następnie podlega klasycznemu przetworzeniu za pomocą warstw połączonych globalnie w decyzję klasyfikacyjną [78].

2.2. Stan pacjentów

Celem zgromadzenia odpowiedniego materiału badawczego, na potrzeby swoich prac autorzy wykonywali serie badań, najczęściej we współpracy z ośrodkami medycznymi. Zazwyczaj w badaniach brała udział grupa osób chorych na chorobę Parkinsona i grupa osób zdrowych jako grupa kontrolna. Osoby zdrowe były dopasowane do grupy osób chorych pod względem płci i wieku. Niektórzy autorzy wyróżniali podgrupy w zależności od wieku pacjentów [26, 44, 48, 49, 72], aby odróżnić zmiany w piśmie wynikające z choroby Parkinsona, od naturalnych zmian wynikających z procesu starzenia. W badaniach mających na celu sprawdzenie możliwości wczesnej diagnostyki choroby Parkinsona na podstawie pisma odręcznego, badani pacjenci byli poddani leczeniu farmakologicznemu celem zminimalizowania łatwo widocznych objawów, tak, aby odpowiadały one osobom we wczesnym stadium rozwoju choroby (tzw. faza „on”). W literaturze obecne są również wyniki badań wykonywanych z udziałem pacjentów w tzw. fazie „off”, czyli w fazie po odstawieniu przepisanej dawki leków neurologicznych. W większości przypadków dotyczą one prób określenia cech pisma osób z widocznymi objawami choroby Parkinsona [26, 47, 54, 55]. W niektórych pracach próbowano wpływać na intensywność objawów podczas akwizycji danych poprzez zastosowanie pomocniczych znaczników wskazujących na docelową wielkość

liter [48]. Badania w fazie „off” wykorzystywane były również do oceny dodatkowych czynników stymulujących proces tworzenia pisma, do których należało równoczesne liczenie [27, 72, 73] lub powtarzanie słów [72].

2.3. Wzorce pisma

Ze względu na łatwość zdefiniowania pojedynczego pociągnięcia w górę lub w dół podczas pisania i jednocześnie stosunkowo prosty algorytm przetwarzania, w wielu badaniach nad analizą pisma osób cierpiących na chorobę Parkinsona jako wzór pisma wybierano ciąg liter połączonych ze sobą i tworzących pętle np. „lllll” [45-51, 73], „lelelele” [42, 74] czy „lililili” [69] lub pojedyncze słowa zawierające litery „l”, „e” lub „h”. Dobrym przykładem takiego słowa jest „hello” lub „hell” [74]. Niestety, nie w każdym języku znajdują się słowa składające się jednocześnie z kilku powyższych liter. Stąd często pacjenci proszeni byli o zapis pseudosłów stworzonych na potrzeby badania, takich jak: „ellehell” [43], „eluhule” i „eludule” [25]. Takie podejście może wprowadzać pewien dyskomfort oraz dodatkowo dekoncentrować osobę badaną, co może wpływać na wynik analizy pisma. Lepszym i bardziej naturalnym rozwiązaniem jest poproszenie pacjentów o zapis pełnego zdania w ojczystym języku. Praktyka taka została przyjęta w niektórych pracach [52, 55, 57, 63, 75], jednak w większości z nich nie do końca wykorzystano jej możliwości z uwagi na fakt, że do dalszej analizy brane były tylko niektóre ze znajdujących się w pisanych zdaniach słowa lub litery [52, 55, 75]. Na przykład w pracy [75], jako wzór pisma wybrano zdanie: „Ein helles grelles Licht” (z niem. *Jasne oślepiające światło*), z którego dalszemu przetwarzaniu zostały poddane tylko sekwencje „ll” ze słów „helles” i „grelles”. Wzorce pisma wykorzystywane w badaniach nad obecnością choroby Parkinsona zestawiono w tabeli 2.1.

Tabela 2.1. Zestawienie stosowanych w badaniach nad chorobą Parkinsona wzorców pisma

WZORCE PISMA				
ZDANIA	SŁOWA	LITERY/SYLABY		INNE
		ROZDZIELONE	POŁĄCZONE	
<i>They wrote eluhule</i> [25] <i>They wrote ellehell at the Motor Behavior Lab of the University of Wisconsin</i> [43, 44] <i>En liesje leerde loesje lopen</i> [52] <i>Veel te veel felle schelle zon</i> [54] <i>Tramvaj dnes už nepojede</i> [57-61] <i>El abecedario es a, b, c, ..., z</i> [63] <i>Ein helles grelles Licht</i> [75] <i>The leveler leveled all levels</i> [72]	<i>eluhule</i> [25] <i>eludule</i> [25] <i>ellehell</i> [43, 44] <i>elle</i> [56] <i>lektorka</i> [57-61] <i>porovant</i> [57-61] <i>nepopadnout</i> [57-61] <i>hell</i> [74] <i>level</i> [72]	<i>e e e e e e e</i> [55] <i>l l l l l l l</i> [57-61] <i>le le</i> [57-61] <i>les les</i> [57-61] <i>b b b b b b b</i> [71] <i>d d d d d d d</i> [71] <i>bd bd bd bd</i> [71]	<i>lllllll</i> [27, 45-51, 56, 62, 73] <i>mmmmmm</i> [62] <i>elelelelelel</i> [54] <i>lelelelelele</i> [74] <i>lililililili</i> [50]	<i>imię</i> [53, 62, 71] <i>adres</i> [53] <i>cyfry od 0 do 9</i> [62] <i>numer ID</i> [62] <i>podpis</i> [62] <i>alfabet</i> [62]

Badanie całych zdań ma tę zaletę, że umożliwia analizę nie tylko procesu tworzenia pisma widocznego na kartce, ale również momentów, w których pióro jest uniesione nad powierzchnią tabletu, między kolejnymi pociągnięciami [57, 59-61, 65, 68]. Ocena tych ruchów może wskazywać dodatkowe cechy diagnostyczne charakterystyczne dla choroby Parkinsona [57]. Innymi wykorzystanymi w badaniach naturalnymi wzorcami pisma są np. imię [53, 62], adres [53], cyfry od 0 do 9 czy alfabet [62].

W 2016 roku autorzy pracy [59] na potrzeby swoich badań nad zmianami w piśmie odręcznym osób chorych na chorobę Parkinsona, utworzyli i opublikowali do użytku zewnętrznego bazę danych o nazwie PaHaW. Baza danych powstała we współpracy Centrum Zaburzeń Ruchu i Kliniki Neurologii Uniwersytetu Masaryka ze Szpitalem Uniwersyteckim św. Anny w Brnie w Czechach. Baza danych zawiera próbki pisma 37 pacjentów z chorobą Parkinsona (19 mężczyzn, 18 kobiet) i 38 osób zdrowych (20 mężczyzn, 18 kobiet). Chorych pacjentów badano bez wykluczenia ich codziennej dawki leków dopaminergicznych, czyli w fazie „on”. Podczas badania pacjenci mieli do wykonania 8 zadań piśmienniczych: rysunek spirali, zapis pojedynczych liter, sylab, słów oraz całego zdania w języku czeskim. Wybór wykorzystanych w badaniu słów: *lektorka* (z cs. *nauczycielka*), *porovnat* (z cs. *porównać*) i *nepopadnout* (z cs. *nie złapać*) autorzy argumentują prostą ortografią oraz tym, że wszystkie te wyrazy można zapisać w sposób ciągły tj. bez konieczności oderwania pióra od powierzchni kartki. Na rysunku 2.1 zamieszczono szablon na podstawie którego pacjenci wykonywali zadania.



Rys. 2.1. Wzór zadań piśmienniczych bazy danych PaHaW

Jako narzędzie do akwizycji próbek pisma użyto tablet graficzny, który umożliwiał rejestrację danych, takich jak: współrzędne X, Y położenia rysika na tablecie, nacisk pióra na powierzchnię tabletu i kąt nachylenia pióra do powierzchni tabletu. Autorzy bazy zdecydowali się na jej publiczne udostępnienie w postaci surowych danych pochodzących bezpośrednio

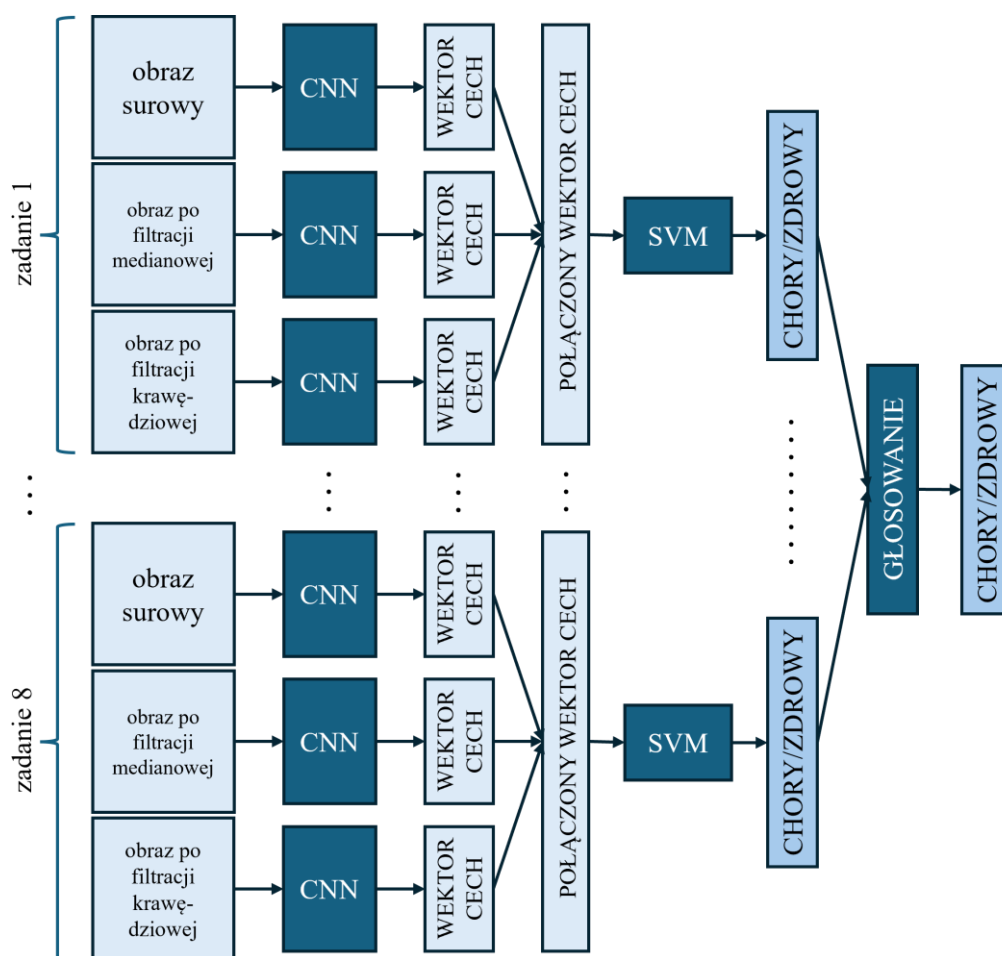
z tabletu. Dzięki możliwości samodzielnego przetwarzania i analizy tych danych, większość autorów późniejszych prac wykorzystwała tę bazę w swoich badaniach, próbując poprawiać wyniki klasyfikacji poprzedników poprzez zastosowanie odmiennych metod przetwarzania i eksploracji danych [60, 64-68, 76, 77]. Również większość badań nad pismem badających potencjał zastosowania konwolucyjnych sieci neuronowych w procesie automatycznej generacji cech i klasyfikacji opiera się na tej bazie [79-83].

2.4. Konwolucyjne sieci neuronowe – aktualny kierunek badań

W najnowszych badaniach, wykonywanych pod kątem diagnozowania choroby Parkinsona na podstawie próbek pisma, wykorzystuje się konwolucyjne sieci neuronowe, które stanowią ważny element intensywnie rozwijającej się obecnie dziedziny jaką jest głębokie uczenie (ang. *deep learning*). Motywacją do ich zastosowania w powyższym problemie jest fakt, że bazują one bezpośrednio na całych obrazach, bez potrzeby ich przetwarzania w poszukiwaniu wektora cech, przez co wydają się być odpowiednim narzędziem do analizy obrazów pisma odręcznego. Dodatkowo cyfrowa rejestracja próbek pisma, umożliwia różnorodną wizualizację tego pisma na drodze wstępnego przetwarzania, celem wyeksponowania na obrazie dodatkowych jego cech, mogących wspomóc sieci konwolucyjne w pozyskiwaniu informacji diagnostycznej z obrazów.

W ciągu ostatnich pięciu lat pojawiło się kilka artykułów pokazujących zastosowanie konwolucyjnych sieci neuronowych w analizie próbek pisma w zadaniu diagnozowania choroby Parkinsona na ich podstawie. We wszystkich omówionych w tym podrozdziale pracach wykorzystano dane z opisanej w rozdziale 2.3. ogólnodostępnej bazy danych PaHaW. Ze względu na małą liczbę danych wejściowych, autorzy prac korzystali z metod *transfer learningu*, polegającego na użyciu wstępnie wytrenowanych sieci i zastosowaniu ich, po douczeniu na nowym zbiorze danych, do rozwiązywania odmiennego problemu, który w tym przypadku stanowi rozpoznawanie choroby Parkinsona na podstawie próbek pisma.

W pracy [79] autorzy wykorzystali wstępnie nauczoną, na zbiorze danych ImageNet, strukturę sieci AlexNet do automatycznej generacji cech. Wektor cech opracowany przez sieć w warstwie w pełni połączonej *fc7*, został następnie podany na klasyfikator typu SVM, celem wypracowania ostatecznej decyzji klasyfikacyjnej. Surowe obrazy prezentujące każde z 8 zadań piśmienniczych zostały poddane transformacjom nieliniowym, celem uzyskania dodatkowych reprezentacji. Trzy reprezentacje danych wejściowych: obraz surowy (D_r), obraz powstały w wyniku filtracji medianowej (D_m) i obraz powstały po filtracji krawędziowej (D_e), posłużyły do uczenia trzech niezależnych sieci dla każdego z ośmiu zadań wykonywanych przez osobę badaną. Cechy wyodrębnione przez trzy sieci były następnie łączone w wektor cech. Koncepcję systemu decyzyjnego wykorzystanego w pracy [79], przedstawiono na rysunku 2.2. Dzięki zastosowaniu odrębnych sieci, możliwe było zbadanie wpływu poszczególnych zadań piśmienniczych oraz ich reprezentacji graficznej na ostateczny wynik klasyfikacji.



Rys. 2.2. Koncepcja systemu decyzyjnego zaproponowanego w pracy [79]

Wyniki dokładności rozpoznania choroby, dla różnych reprezentacji danych na obrazie i zadań piśmienniczych, otrzymane w pracy [79] zamieszczono w tabeli 2.2. Najlepsze rezultaty, niezależnie od reprezentacji danych, uzyskano generując wektory cech na podstawie wszystkich zadań piśmienniczych. Należy jednak zwrócić uwagę na fakt, że wynik ten jest nieznacznie lepszy, niż w przypadku samej spirali, co może oznaczać, że to właśnie ten wzór ma największy wpływ na ostateczny wynik. Biorąc pod uwagę tylko wzorce pisma, najlepszą dokładność autorzy uzyskali dla słowa *nepopadnout* i wynosił on 68 %. Natomiast wyniki dokładności klasyfikacji uzyskane w przypadku zdania, wynoszą ok. 50 %, co jest wysoce niezadowalającym wynikiem.

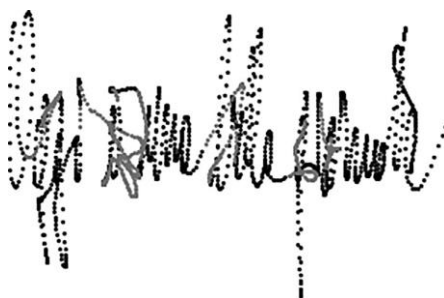
Tabela 2.2. Zestawienie wyników dokładności rozpoznania choroby Parkinsona w pracy [79]

	Dr	Dm	De	Dr + Dm	Dr + De	Dm + De	Dr+Dm+De
<i>spirala</i>	0,57	0,65	0,65	0,67	0,70	0,65	0,76
<i>lll</i>	0,53	0,55	0,57	0,55	0,52	0,50	0,62
<i>le le le</i>	0,48	0,51	0,54	0,51	0,52	0,55	0,57
<i>les les</i>	0,50	0,57	0,55	0,54	0,52	0,57	0,60
<i>lektorka</i>	0,49	0,58	0,52	0,54	0,52	0,51	0,60

Tabela 2.2. Zestawienie wyników dokładności rozpoznania choroby Parkinsona w pracy [79] (cd.)

	Dr	Dm	De	Dr + Dm	Dr + De	Dm + De	Dr+Dm+De
porovant	0,46	0,49	0,48	0,50	0,49	0,47	0,51
nepopadnout	0,54	0,64	0,60	0,65	0,59	0,63	0,68
zdanie	0,48	0,49	0,48	0,50	0,49	0,49	0,51
razem	0,58	0,68	0,66	0,73	0,76	0,79	0,83

W pracy [80] zastosowano identyczną koncepcję systemu decyzyjnego jak w pracy [79]. Dla każdego zadania piśmienniczego generowano trzy reprezentacje graficzne podane na wejście trzech niezależnych sieci. Uzyskane w ten sposób wektory cech podawano na klasyfikator. W tym badaniu wykorzystano jednak inną strukturę sieci – VGG16, która zawiera większą liczbę warstw konwolucyjnych niż sieć AlexNet użyta przez autorów pracy [79]. Autorzy pracy [80] zaproponowali również zupełnie inną metodę prezentacji danych na obrazie. W przeciwieństwie do autorów [79], rekonstrukcji pisma nie dokonano poprzez liniowe połączenie zarejestrowanych punktów, ale poprzez wykreślenie samych punktów. Do tego uwzględniono położenie pióra zarówno na powierzchni (kolor czarny) jak i w powietrzu (kolor szary). Ideę wizualizacji danych na obrazie prezentuje rysunek 2.3.



Rys.2.3. Prezentacja danych na obrazie zastosowana przez autorów [80]

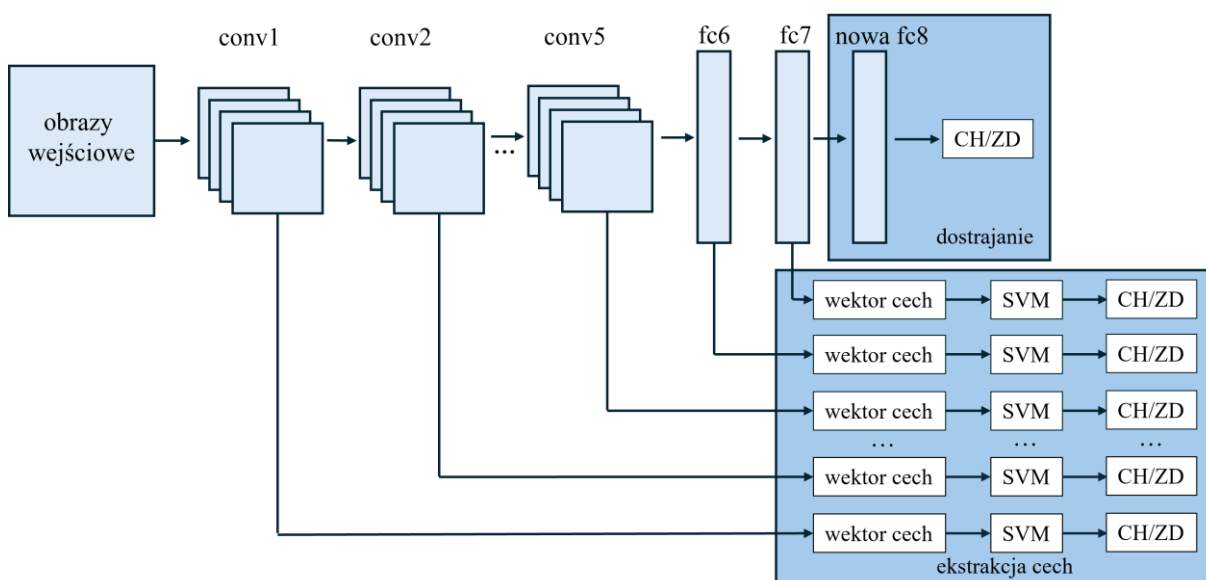
Autorzy twierdzą, że w ten sposób wprowadzają do obrazu pewną informację dynamiczną, ponieważ większa prędkość pisania spowoduje, że punkty będą bardziej odległe od siebie. W pracy przetestowano kilka różnych klasyfikatorów m.in. SVM, losowy las czy AdaBoost. Nie rozpatrywano tutaj natomiast przypadków zastosowania pojedynczych reprezentacji lub ich kombinacji. Wszystkie wyniki podane w pracy odnoszą się do wektorów cech uzyskanych na podstawie trzech reprezentacji. Wyniki dokładności rozpoznania choroby Parkinsona przy pomocy klasyfikatora SVM, uzyskane w pracy [80] zamieszczono w tabeli 2.3.

Tabela 2.3. Zestawienie wyników dokładności w pracy [80] podane w procentach

	D_r+D_m+D_e
spirala	75,00
lll	58,33
le le le	53,75
les les	57,08
lektorka	75,41
porovant	63,75
nepopadnout	57,50
zdanie	67,08
razem	86,67

Mimo wykorzystania bardziej złożonej struktury sieci i innej metody reprezentacji danych na obrazie niż w pracy [79], uzyskane dokładności rozpoznania są na podobnym poziomie. Otrzymano nieznacznie lepsze wyniki dla niektórych zadań piśmienniczych, np. dla słów i zdania. Najwyższy wynik dokładności wynoszący 75,41 % w przypadku wzorca pisma, autorzy uzyskali dla słowa *lektorka*.

W pracy [81] wykorzystano sieci AlexNet, wstępnie przeszkolone na dwóch bazach danych uczących: ImageNet i MNIST, do klasyfikacji choroby Parkinsona także na podstawie obrazów pisma pochodzących z bazy PaHaW. Zbadano dwa powszechne podejścia do *transfer learningu*: dostrajanie (ang. *fine-tuning*) i ekstrakcję cech (ang. *feature extraction*). W drugim przypadku, wykorzystano wstępnie przeszkoloną sieć do generacji wektora cech, który następnie podano na klasyfikator SVM podobnie jak w pracach [79, 80]. Najlepszy rezultat uzyskano wyodrębniając cechy z piątej warstwy konwolucyjnej *conv5*. W podejściu opartym na dostrajaniu, użyto wstępnie przeszkoloną architekturę sieci, zastępując ostatnią w pełni połączoną warstwę nową, której liczba neuronów odpowiada liczbie klas w docelowym zadaniu klasyfikacji. Koncepcję badań zaproponowanych w pracy [81] zamieszczono na rysunku 2.4.



Rys. 2.4. Zaproponowany w pracy [81] system do rozpoznawania choroby Parkinsona

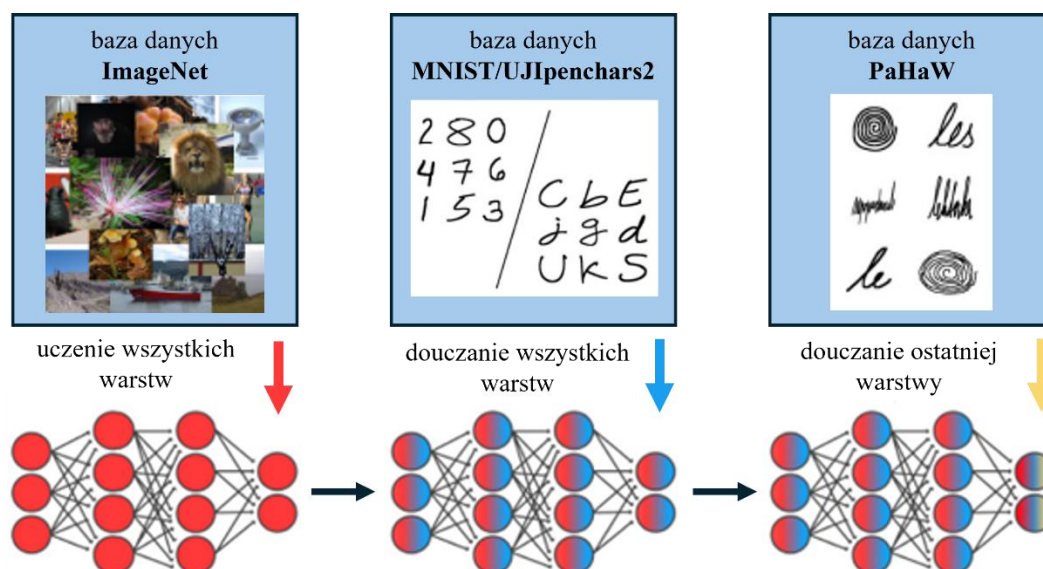
W celu zwiększenia liczby danych uczących przeprowadzono tzw. augmentację danych, polegającą na wprowadzeniu do materiału treningowego zmodyfikowanych kopii istniejących danych. W ramach augmentacji przeprowadza się szereg operacji na pojedynczym obrazie generując w ten sposób nowe dane. Ważne jest jednak, aby powstałe dane w sposób autentyczny odzwierciedlały rzeczywiste przykłady. W omawianej pracy natomiast przeprowadzono między innymi operację obrotu obrazu o zadany kąt (45° , 90° , 135° , 180° , 225° , 270° , 315°) oraz przerzucenie w pionie i w poziomie. Takie podejście w odniesieniu do próbek pisma nie jest właściwe, gdyż generuje dane niebędące autentycznymi obrazami pisma np. odbicie lustrzane lub obrót o 180° . Zastosowanie sieci wytrenowanej na takich wzorcach

do testowania rzeczywistych próbek pisma może nie być efektywne. W omawianym przypadku, zbiór danych testujących został wyodrębniony spośród obrazów uzyskanych na drodze augmentacji. Do tego z pracy jednoznacznie nie wynika, czy obrazy pochodzące od tej samej osoby nie znajdują się jednocześnie w zbiorze testowym i treningowym. Może to być przyczyną uzyskania zaskakująco wysokich wyników dokładności w porównaniu z poprzednimi pracami [79, 80], zwłaszcza w przypadku podejścia bazującego na generacji wektora cech za pomocą sieci konwolucyjnej. Wyniki dokładności rozpoznania choroby, uzyskane w pracy [81] zamieszczono w tabeli 2.4. Analizując uzyskane przez autorów wyniki, trudno jednoznacznie stwierdzić, które podejście, ekstrakcja cech czy dostrajanie, pozwala uzyskać lepsze wyniki oraz która źródłowa baza danych jest korzystniejsza.

Tabela 2.4. Zestawienie wyników dokładności rozpoznania choroby Parkinsona w pracy [81] podane w procentach

	ImageNet wektor cech z <i>conv5</i>	MNIST wektor cech z <i>conv5</i>	ImageNet dostrajanie	MNIST dostrajanie
<i>spirala</i>	93,10	94,97	98,28	80,17
<i>111</i>	76,72	86,21	91,38	70,69
<i>le le le</i>	79,49	86,32	96,58	71,79
<i>les les</i>	73,21	77,68	75,89	63,39
<i>lektorka</i>	72,32	81,25	83,93	59,82
<i>porovant</i>	89,29	68,75	75,00	62,50
<i>nepopadnout</i>	78,99	83,19	83,19	68,91
<i>zdanie</i>	77,48	86,49	54,95	50,45
razem	87,05	74,53	84,19	77,28

W [82] badano możliwości wielokrotnego dostrajania struktur sieci VGG16 i SqueezeNet za pomocą różnych baz danych. W pierwszym kroku wszystkie warstwy sieci trenowane były na zbiorze danych ImageNet. W następnym kroku wszystkie warstwy douczane były za pomocą baz danych zawierających obrazy pisanych odręcznie cyfr (baza MNIST) lub liter (baza UIIPenchars2). Ostatnim etapem było dostrojenie ostatnich warstw sieci za pomocą danych z bazy PaHaW z wykluczeniem obrazów prezentujących całe zdanie. Wykorzystaną koncepcję wielokrotnego dostrajania zamieszczono na rysunku 2.5. Wstępne przetwarzanie danych, podobnie jak w [81], obejmowało augmentację. W tym przypadku jednak autorzy zdecydowali się jedynie na obrót próbek pisma o 20° , co reprezentuje naturalne pochylenie pisma. Uzyskane wyniki dokładności rozpoznania choroby dla różnych zestawów danych uczących, zaprezentowano w tabeli 2.5 i 2.6. Autorom poprzez wielokrotne douczanie sieci udało się w większości przypadków uzyskać wyniki lepsze od autorów [79, 80]. W prawie wszystkich przypadkach zastosowanie wielokrotnego dostrajania przewyższa konwencjonalnie dostrajane sieci. W większości przypadków korzystniejsze wyniki otrzymano dla architektury VGG16 niż dla architektury SqueezeNet. Najwyższy wynik dokładności rozpoznania choroby w przypadku pisma uzyskano podobnie jak w [79], dla słowa *nepopadnout* i wynosił on 78,5 %.



Rys. 2.5. Koncepcja wielokrotnego dostrajania wykorzystana w publikacji [82]

Tabela 2.5. Zestawienie wyników dokładności rozpoznania choroby Parkinsona przez sieć VGG16 w pracy [82] podane w procentach

	ImageNet	ImageNet UJIPenchars2	ImageNet MNIST	MNIST UJIPenchars2	MNIST	ImageNet MNIST UJIPenchars2
<i>spirala</i>	78,5	81,6	83,0	85,4	79,5	85,8
<i>lll</i>	64,5	67,6	66,8	65,0	65,0	68,0
<i>le le le</i>	73,5	71,3	68,0	64,9	65,1	74,7
<i>les les</i>	70,8	69,9	70,8	68,6	68,5	72,7
<i>lektorka</i>	72,7	74,7	73,4	72,2	68,4	76,1
<i>porovant</i>	68,1	67,8	68,7	68,5	64,7	76,0
<i>nepopadnout</i>	75,8	78,4	78,5	77,3	72,1	78,5

Tabela 2.6. Zestawienie wyników dokładności rozpoznania choroby Parkinsona przez sieć SqueezeNet w pracy [82] podane w procentach

	ImageNet	ImageNet UJIPenchars2	ImageNet MNIST	MNIST UJIPenchars2	MNIST	ImageNet MNIST UJIPenchars2
<i>spirala</i>	76,9	79,8	77,1	75,3	73,5	85,7
<i>lll</i>	63,9	64,6	64,3	65,1	64,7	66,3
<i>le le le</i>	66,9	67,2	66,9	66,7	66,6	71,4
<i>les les</i>	66,8	67,9	67,0	67,5	67,0	68,7
<i>lektorka</i>	65,2	70,6	63,3	73,4	69,3	71,9
<i>porovant</i>	64,6	65,8	66,9	66,6	64,2	69,4
<i>nepopadnout</i>	70,4	71,8	75,0	72,0	71,6	76,1

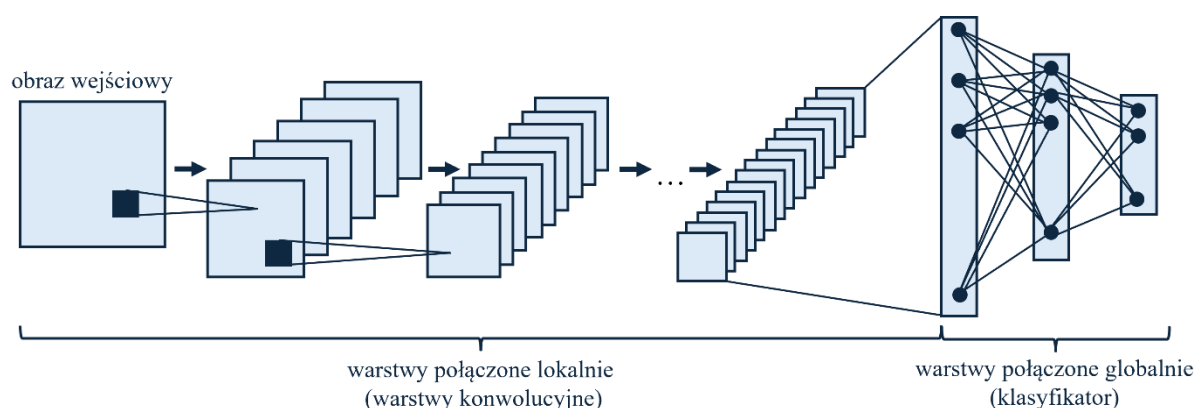
Omówione w tym rozdziale prace w sposób znaczący ukształtowały podejście badawcze, a otrzymane przez autorów wyniki stały się odniesieniem do wyników uzyskanych w ramach tej rozprawy. W kontekście niniejszej dysertacji należy w szczególności zwrócić uwagę na

wyniki otrzymane dla całego zdania, gdyż to właśnie one będą służyły jako porównanie dla wyników uzyskanych przez autora.

3. Konwolucyjne sieci neuronowe – charakterystyka narzędzia

Konwolucyjne sieci neuronowe CNN (ang. *Convolutional Neural Network*) to rodzaj sieci głębokich obecnie powszechnie stosowanych w przetwarzaniu obrazów. W rzeczywistości sieci te stanowią połączenie nienadzorowanego procesu automatycznej generacji cech diagnostycznych obrazu poprzez zastosowanie wielu ukrytych warstw konwolucyjnych z końcowym klasyfikatorem [84]. W strukturze CNN można wyróżnić zatem dwie części (rysunek 3.1):

- Szereg kolejnych warstw konwolucyjnych (dostarczających na końcu zbiór cech diagnostycznych), których neurony są połączone lokalnie z niewielkim obszarem danych wejściowych dla danej warstwy.
- Część końcowa sieci o pełnym połączeniu neuronów, stanowiąca właściwy klasyfikator.

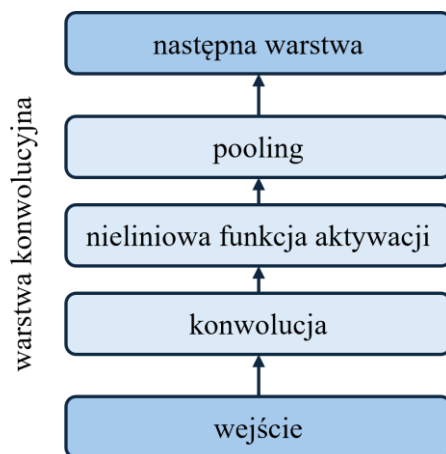


Rys. 3.1. Struktura sieci CNN

Na wejście sieci podawany jest obraz oryginalny w postaci macierzy zawierającej stopnie jasności pikseli w przypadku obrazów monochromatycznych lub w przypadku obrazów kolorowych w postaci tensora składającego się z trzech kanałów, reprezentujących kolejno komponenty odpowiadające barwom: R – czerwony, G – zielony, B – niebieski [85]. Pierwsze warstwy mają za zadanie wykrywać pewne prymitywne elementy obrazu np. krawędzie czy charakterystyczne skupiska pikseli. Kolejne warstwy wykonując podobne operacje na tensorze wyjściowym warstwy poprzedzającej, rozpoznają coraz większe elementy obrazów. W wyniku takiego przetwarzania ostatnia warstwa konwolucyjna zawiera zintegrowane cechy procesu, które mogą stanowić atrybuty wejściowe dla dowolnego klasyfikatora [84] lub być przetwarzane w sygnały wyjściowe CNN przez szereg warstw w pełni połączonych

zakończonych zazwyczaj warstwą *softmax*. Poszczególne warstwy konwolucyjne, z których składa się sieć CNN, zawierają zazwyczaj 3 poziomy (rysunek 3.2):

- poziom właściwej liniowej konwolucji (filtracji),
- poziom nieliniowej funkcji aktywacji,
- poziom redukcji wymiaru obrazu (pooling)



Rys. 3.2. Składniki typowej warstwy konwolucyjnej [85]

3.1. Operacja splotu

Nazwa konwolucyjnych sieci neuronowych pochodzi od operacji konwolucji (splotu), która stanowi podstawę działania tych sieci. W przypadku sygnałów dyskretnych operację splotu można wyrazić wzorem 3.1 [85]:

$$y(n) = x(n) * w(n) = \sum_{i=-\infty}^{\infty} x(i)w(n-i) \quad (3.1)$$

gdzie: $x(n)$ – sygnał wejściowy

$w(n)$ – jądro przekształcenia

$y(n)$ – sygnał wyjściowy, będący odwzorowaniem (mapą) cech procesu

Jeżeli dane wejściowe mają postać obrazu, to jest on reprezentowany przez macierz $\mathbf{I}$, której elementy odpowiadają stopniu jasności pikseli $I(m,n)$. Operacja splotu dwuwymiarowego wymaga zatem użycia dwuwymiarowego jądra przekształcenia $\mathbf{K}$, nazywanego również filtrem lub maską filtrującą. W trakcie wykonywania operacji konwolucji, maska filtrująca przemieszcza się po całym obrazie z określonym krokiem (ang. *stride*), generując kolejne wartości pikseli obrazu wynikowego z uwzględnieniem wartości maski filtrującej, zgodnie ze wzorem 3.2 [85]:

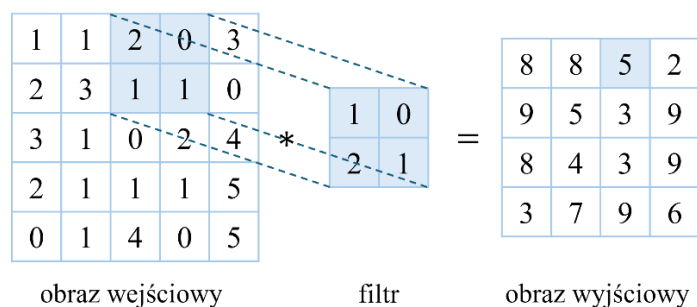
$$Y(m,n) = I(m,n) * K(m,n) = \sum_i \sum_j I(i,j)K(m-i,n-j) \quad (3.2)$$

Zgodnie ze wzorem 3.2, w klasycznej filtracji konwolucyjnej, filtr zostaje odwrócony względem obrazu wejściowego (wraz ze wzrostem i rośnie indeks obrazu wejściowego, a maleje indeks filtru). W praktycznych implementacjach często stosuje się jednak operację,

w której jądro nie jest odwracane, zgodnie ze wzorem 3.3. Takie podejście jest często stosowane w bibliotekach sieci neuronowych, ponieważ upraszcza implementację i interpretację działania filtrów.

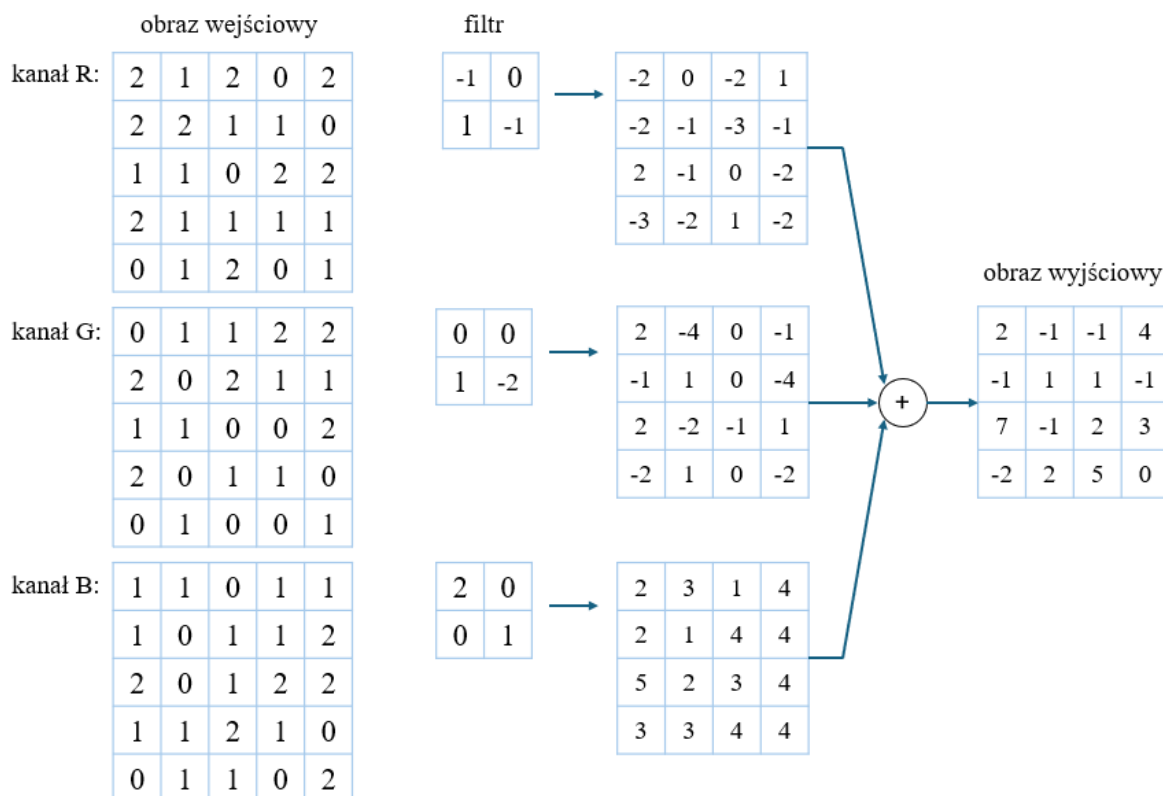
$$Y(m, n) = I(m, n) * K(m, n) = \sum_i \sum_j I(m + i, n + j) K(i, j) \quad (3.3)$$

Przykład operacji konwolucyjnej wykonywanej przez sieci neuronowe dla kroku przesunięcia filtru równego 1 przedstawiono na rysunku 3.3.



Rys. 3.3. Przykład operacji wykonywanej w warstwie konwolucyjnej sieci

W przypadku obrazów kolorowych, w pierwszej warstwie sieci CNN każdy z trzech kanałów RGB obrazu jest filtrowany przez filtr, składający się z wag dla każdego z tych kanałów. Wyniki filtracji dla każdego kanału są sumowane, tworząc obraz wynikowy [84]. Przykład operacji konwolucji dla danych w postaci tensora przedstawiono na rysunku 3.4.



Rys. 3.4. Przykład konwolucji obrazu reprezentowanego przez kanały RGB

Operację konwolucji przeprowadza się wielokrotnie, wykorzystując filtry o różnych wartościach wag. Liczba powstałych w wyniku konwolucji obrazów wyjściowych jest równa liczbie zastosowanych filtrów, a zbiór tych obrazów wyjściowych tworzy tensor nazywany mapą cech (ang. *feature map*). W kolejnych warstwach konwolucyjnych filtr, realizujący operację konwolucji, jest również wielowymiarowy, przy czym liczba kanałów filtru jest równa liczbie obrazów wejściowych, które podlegają przetworzeniu w danej warstwie [84].

Na pojedynczą warstwę konwolucyjną można spojrzeć ekwiwalentnie jako na zestaw neuronów, z których każdy, za pomocą swojego pola recepcyjnego (maski filtru), analizuje tylko fragment obrazu. W związku z tym operacja konwolucji charakteryzuje się trzema ważnymi zaletami w porównaniu do zwykłych operacji macierzowych w sieciach klasycznych. Są to: rzadkie oddziaływanie (ang. *sparse interactions*), współdzielenie parametrów (ang. *parameter sharing*) i ekwiwariancja względem przesunięcia (ang. *equivariant representations*) [84, 85].

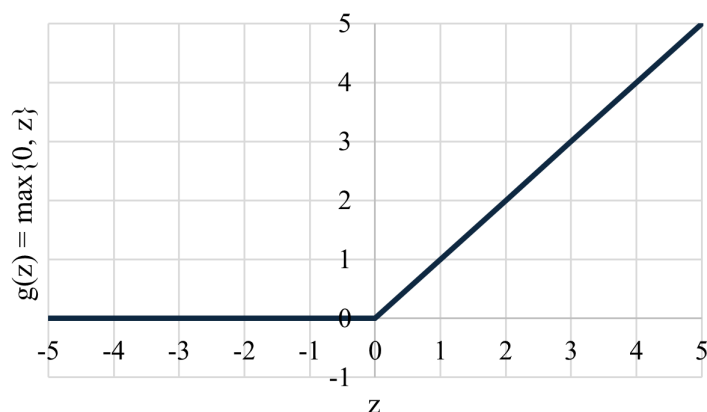
Rzadkie oddziaływania występujące w sieciach CNN, wynikają z lokalności połączeń. Neuron nie jest połączony ze wszystkimi sygnałami wejściowymi, a jedynie z niewielką liczbą pikseli znajdujących się w zasięgu działania maski filtrującej. Przekłada się to na mniejszą liczbę parametrów do przechowywania, co wpływa na zmniejszenie wymagań pamięciowych systemu i podnosi efektywność.

Współdzielenie parametrów oznacza stosowanie tego samego zestawu wag filtru, tworzącego jeden obraz wyjściowy. Takie podejście w sposób znaczny zmniejsza złożoność obliczeniową procesu uczenia, gdyż zamiast adaptacji parametrów wielu masek filtrujących różne obszary obrazu, wystarczy dobór parametrów jednej wspólnej maski przeznaczonej do tworzenia jednego obrazu wyjściowego.

Ekwiwariancja względem przesunięcia oznacza, że jeżeli wejście ulegnie zmianie, to wynik wyjściowy zmieni się dokładnie w ten sam sposób. W szczególności funkcja $f(x)$ jest ekwiwariantna względem funkcji g , jeżeli $f(g(x))=g(f(x))$. Na przykład, jeżeli obraz przesunięty o jeden piksel poddany zostanie operacji splotu, to wynik będzie taki sam, jak gdyby zastosować splot na obrazie oryginalnym i przesunąć obraz wynikowy o jeden piksel.

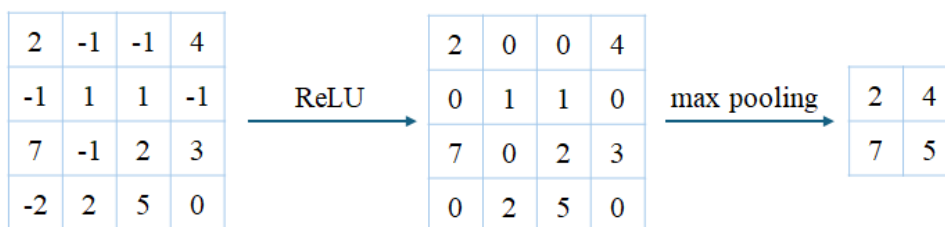
3.2. Nieliniowa funkcja aktywacji ReLu i pooling

Sumacyjny sygnał wyjściowy konwolucji poddawany jest działaniu funkcji aktywacji. Funkcją najczęściej wykorzystywaną w większości jednokierunkowych sieci neuronowych jest przedstawiona na rysunku 3.5, funkcja ReLU (ang. *rectified linear unit*). Powstały obraz wynikowy poddaje się działaniu operacji statystycznego łączenia sąsiednich wyników tzw. *pooling* z wykorzystaniem maski filtrującej przesuwającej się po obrazie. Najczęściej spotykanymi rodzajami poolingu jest *max pooling* i *average pooling*, zwracające kolejno wartość maksymalną i wartość średnią wyników neuronów znajdujących się w obszarze maski [84].



Rys. 3.5. Funkcja aktywacji ReLu

Obraz wynikowy otrzymany po zastosowaniu funkcji aktywacji ReLu i operacji *max pooling* z krokiem równym 2 na obrazie wyjściowym z rysunku 3.4 pokazano na rysunku 3.6.

Rys. 3.6. Przykład działania na obrazie funkcji ReLu i operacji *max pooling*

Operacja *pooling* prowadzi do redukcji wymiaru obrazu poprzez uogólnienie cech reprezentowanych przez sąsiednie grupy pikseli. Jednocześnie powoduje, że sieć toleruje niewielkie przesunięcie obrazu (ang. *location invariance*) [86]. Chcąc zachować odpowiedni wymiar obrazów wyjściowych lub nie dopuścić do zbyt gwałtownego zmniejszenia wymiaru w kolejnych warstwach, stosuje się tzw. *zero padding* polegający na uzupełnieniu brzegów obrazu zerami [84]. Zazwyczaj jest to uzupełnienie symetryczne z każdej strony obrazu. Przykładowy obraz uzupełniony pojedynczymi liniami zer zaprezentowano na rysunku 3.7.

0	0	0	0	0	0	0
0	1	1	2	0	3	0
0	2	3	1	1	0	0
0	3	1	0	2	4	0
0	2	1	1	1	5	0
0	0	1	4	0	5	0
0	0	0	0	0	0	0

Rys. 3.7. Przykład zastosowania *zero padding* na obrazie

3.3. Warstwy w pełni połączone - Fully Connected Layers

W kolejnych konwolucyjnych warstwach sieci zmniejsza się wymiar obrazów, ale zwiększa się ich liczba, czyli głębokość tensora. Macierze reprezentujące obrazy ostatniej warstwy konwolucyjnej są przetwarzane na drodze konkatencji na wektory stanowiące dane wejściowe dla sieci w pełni połączonej, gdzie każdy neuron jest połączony z każdym sygnałem warstwy poprzedniej. Liczba neuronów w ostatniej warstwie pełnego połączenia jest równa liczbie klas, a sygnał wyjściowy i -tego neuronu określony jest wzorem 3.4 [84]:

$$u_i(x) = \sum_j w_{ij} x_j + w_0 \quad (3.4)$$

gdzie: w_{ij} – wagi neuronu podlegające adaptacji w procesie uczenia przez optymalizację funkcji celu

w_0 – wyraz wolny (bias) dla i -tego neuronu

x_j – sygnały wejściowe

W sieciach CNN najczęściej na wyjściu stosuje się warstwę typu *softmax*, gdzie sygnały wyjściowe ostatniej warstwy pełnego połączenia są bezpośrednio przetwarzane na prawdopodobieństwo przynależności do określonej klasy [84]. Prawdopodobieństwo przynależności wektora x do i -tej klasy ($i = 1, 2, \dots, K$) obliczane jest ze wzoru 3.5:

$$\text{softmax}(u_i) = \frac{e^{u_i}}{\sum_{j=1}^K e^{u_j}} \quad (3.5)$$

3.4. Uczenie sieci CNN

Parametry sieci CNN takie jak: liczba warstw konwolucyjnych, liczba neuronów w poszczególnych warstwach (liczba filtrów), wymiary filtrów, krok przesunięcia maski (ang. *stride*) oraz liczba uzupełnień zerami (ang. *zero padding*) są ustalane przez użytkownika podczas opracowywania struktury sieci. Ze względu na ich liczbę, prawidłowy dobór powyższych parametrów wymaga doświadczenia i wiedzy eksperckiej w zakresie budowy sieci CNN. Sam proces uczenia warstw konwolucyjnych sieci polega na doborze wag przesuwających się filtrów. Ponadto, w procesie uczenia sieci CNN dobierane są wagi warstw pełnego połączenia. Zatem liczba parametrów, których wartości są dobierane podczas uczenia zależy od liczby warstw, liczby filtrów i wielkości masek filtrujących w warstwach konwolucyjnych oraz liczby neuronów w warstwach w pełni połączonych. W związku z tym liczba ta może wynosić nawet od kilku do kilkuset milionów parametrów. Na przykład struktura AlexNet, przetwarzająca w pierwszej warstwie obrazy o wymiarze 227×227 za pomocą filtru o wymiarze 11×11 , wymaga doboru 61 milionów parametrów w procesie uczenia. Znalezienie optymalnych wartości tak dużej liczby parametrów wymaga użycia odpowiedniej metody optymalizacyjnej. W sieciach CNN najczęściej stosuje się metody gradientowe, których podstawą działania jest znajomość gradientu funkcji celu. Ze względu na prostotę i stosunkowo

małą złożoność obliczeniową najczęściej wykorzystuje się algorytm największego spadku lub algorytm największego spadku z momentem rozpędowym. Algorytm największego spadku (ang. *gradient descent*) aktualizuje parametry sieci poprzez minimalizację funkcji celu, w kierunku jej największego spadku (przeciwnym do gradientu) wg wzoru 3.6 [86]:

$$\theta_{k+1} = \theta_k - \eta \nabla_{\theta_k} L(\theta_k) \quad (3.6)$$

gdzie: k – numer iteracji

η – współczynnik uczenia (ang. *learning rate*)

$\nabla_{\theta_k} L(\theta_k)$ – gradient funkcji celu.

Metoda ta jest jednak mało efektywna, ze względu na wolną zbieżność oraz brak postępów minimalizacji w okolicy punktu optymalnego. Lepszą efektywność uczenia zapewnia metoda największego spadku z momentem rozpędowym (ang. *gradient descent with momentum*), w której aktualizacja wag uwzględnia dodatkowy czynnik niezależny od aktualnej wartości gradientu zwany momentem wg wzoru 3.7 [84]:

$$\theta_{k+1} = \theta_k - \eta \nabla_{\theta_k} L(\theta_k) + \gamma(\theta_k - \theta_{k-1}) \quad (3.7)$$

gdzie: γ – współczynnik momentu, przyjmowany z przedziału $[0,1]$

Zastosowanie algorytmu rozpędowego powoduje znaczne przyspieszenie obliczeń, poprzez zwiększenie wartości kroku w pobliżu minimum.

Korzystanie z powyższych metod wymaga jednak przetworzenia całego zbioru danych uczących w każdym kroku metody, co jest wysoce nieefektywne ze względu na ograniczenia sprzętowe i czas wykonywania obliczeń [86]. Aby przyspieszyć proces uczenia stosuje się technikę tzw. mini zbiorów uczących (ang. *mini batch*), polegającą na użyciu w każdej iteracji niewielkiej części zbioru uczącego, losowanego każdorazowo [84]. W takim podejściu gradient liczony na całym zbiorze danych aproksymuje się gradientem liczonym na części tego zbioru (*mini batchu*). Można zapisać za pomocą zależności 3.8 [86]:

$$\nabla_{\theta} \frac{1}{N} \sum_i L(x_i) \approx \nabla_{\theta} \frac{1}{M} \sum_{x \in B} L(x) \quad (3.8)$$

gdzie: N – liczba przykładów w zbiorze treningowym

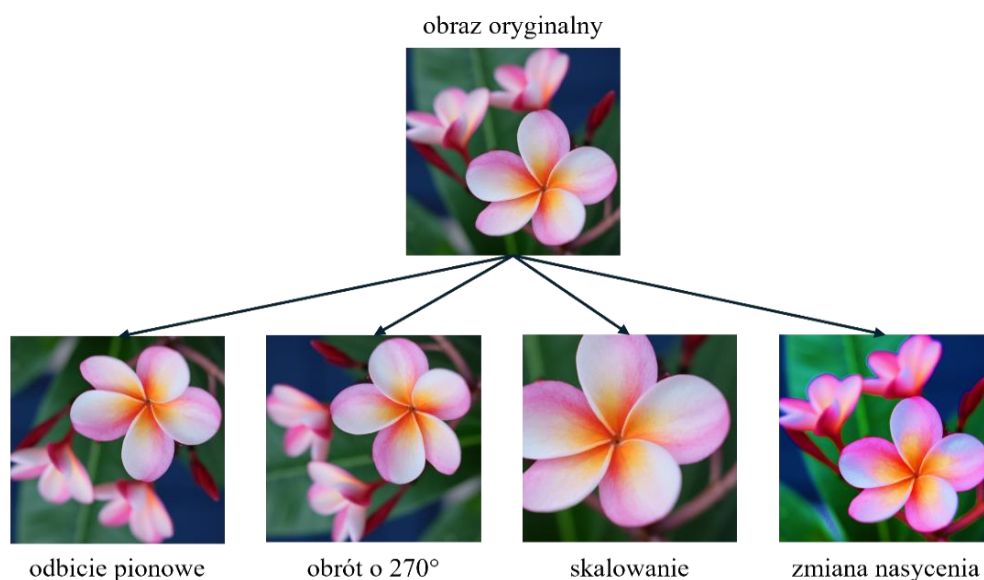
B – wybrany podzbiór (*mini batch*)

M – liczba elementów w wybranym podzbiorze (*mini batchu*)

Metoda ta, w zależności od użytego algorytmu optymalizacyjnego, nosi nazwę stochastycznej metody spadku gradientu SGD (ang. *stochastic gradient descent*) lub stochastycznej metody spadku gradientu z momentem SGDM (ang. *stochastic gradient descent with momentum*). Jej zastosowanie znacząco przyspiesza działanie algorytmu, ale kroki wykonywane w takim podejściu są zaszumione i nie zawsze będą poprawnie minimalizować funkcję kosztu na całym zbiorze [87]. W praktyce okazuje się jednak, że szum w procesie uczenia pełni funkcję regularizacyjną i pozwala poprawić generalizację [86].

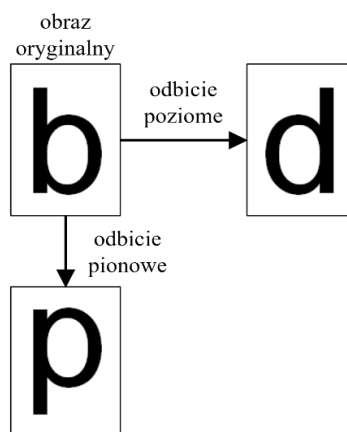
3.5. Augmentacja danych

Aby skutecznie przeprowadzić proces uczenia sieci CNN, potrzebna jest odpowiednio duża liczba danych uczących. Zastosowanie zbyt małego zbioru treningowego w przypadku sieci o dużej liczbie wag będzie prowadzić do nadmiernego dopasowania (ang. *overfitting*). Jednak w praktyce, często liczba posiadanych danych jest ograniczona, a zebranie większego zbioru nie jest proste, a czasem bywa niemożliwe. W takim przypadku przeprowadza się tzw. augmentację danych, polegającą na zwiększeniu zbioru treningowego przez odpowiednie przekształcenia danych wejściowych. Ze względu na metodę przetwarzania można wyróżnić wiele technik augmentacyjnych. W tradycyjnym podejściu wykorzystuje się m.in. transformacje geometryczne (przesunięcie obrazu, skalowanie, obrót czy odbicie) lub transformacje koloru (zmianę jasności, kontrastu czy nasycenia). Obecnie możliwe jest również wygenerowanie nowych, sztucznych obrazów na podstawie danych treningowych np. poprzez zastosowanie sieci GAN (ang. *Generative Adversarial Networks*) [88, 89]. Przykładowe techniki augmentacji zamieszczono na rysunku 3.8.



Rys. 3.8. Przykładowe operacje na obrazie wykonywane w celu augmentacji danych

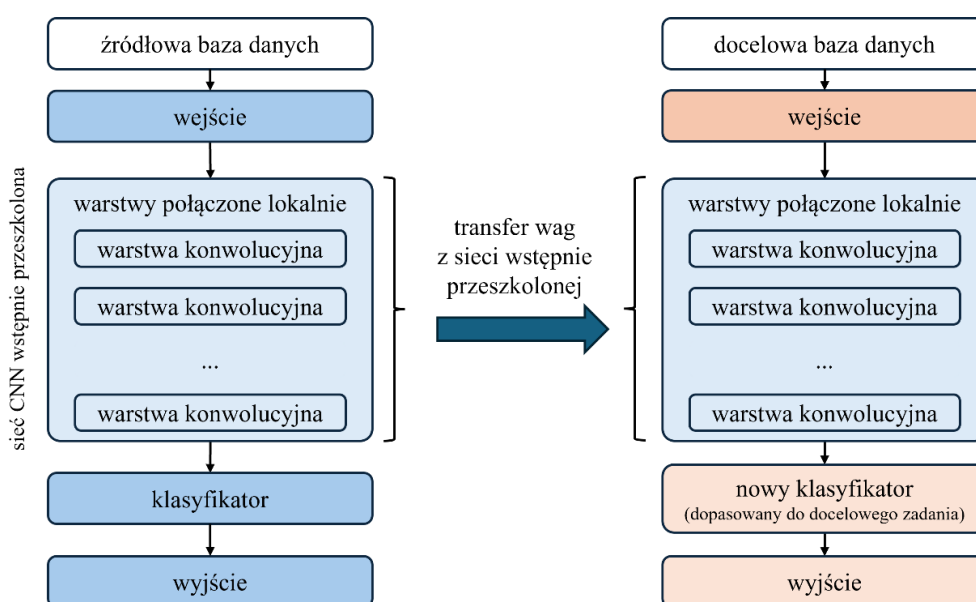
Stosując kilka technik augmentacyjnych można w znaczny sposób zwiększyć liczbę danych treningowych. Jednak należy pamiętać, że obrazy powstałe w wyniku takiego przetwarzania muszą nadal w sposób poprawny reprezentować daną klasę i odpowiednio charakteryzować rodzaj danych. W przypadku niektórych typów obrazów np. pisma odręcznego zastosowanie większości przekształceń augmentacyjnych może prowadzić do utworzenia danych błędnie prezentujących daną klasę lub silnie zdeformowanych w stosunku do rzeczywistych przypadków. Dobrym przykładem jest problem rozpoznawania znaków pisanych. Jak pokazano na rysunku 3.9, litera „b” odbita wzdłuż osi X staje się literą „d”, a odbita wzdłuż osi Y przypomina literę „p”.



Rys. 3.9. Zmiana klasy obiektu pod wpływem operacji odbijania (ang. *flipping*)

3.6. Transfer learning

Konstrukcja własnej wystarczająco rozbudowanej struktury sieci CNN oraz zastosowanie jej do rozwiązywania zadań klasyfikacji wymaga zarówno wiedzy z tego zakresu, jak również użycia odpowiednio dużej bazy wzorców. Natomiast zastosowanie augmentacji nie zawsze okazuje się wystarczające. Obecnie możliwe jest zastosowanie alternatywnego podejścia tzw. *transfer learningu*, polegającego na wykorzystaniu sieci nauczanej do rozwiązywania zupełnie innego zadania jako struktury bazowej i douczenie tej struktury za pomocą nowej bazy danych uczących. Schemat koncepcyjny techniki *transfer learningu* pokazano na rysunku 3.10. Takie podejście jest możliwe ze względu na charakterystyczny sposób przetwarzania danych obrazowych w kolejnych warstwach konwolucyjnych sieci CNN, koncentrujący się nie na całym obrazie, ale na jego prymitywnych elementach [84]. Dzięki takiemu rozwiązaniu możliwe jest ogromne skrócenie czasu uczenia oraz polepszenie zdolności generalizacyjnych sieci.



Rys. 3.10. Schemat koncepcyjny *transfer learningu*

Istnieją dwie możliwości zastosowania *transfer learningu*: ekstrakcja cech (ang. *feature extraction*) oraz dostrajanie (ang. *fine-tuning*). W pierwszym przypadku wykorzystuje się wstępnie wytrenowaną strukturę sieci do generacji wektora cech, którym następnie zasilony zostaje dowolny klasyfikator. Zestaw cech może zostać wyodrębniony z dowolnej warstwy wstępnie przeszkolonej sieci, ale najbardziej powszechną strategią jest użycie wektora pochodzącego z wybranej warstwy w pełni połączonej. Podejście ekstrakcji cech nie zawsze wymaga douczania struktury sieci, a jedynie wykorzystania informacji nabytej w trakcie wstępnego szkolenia, przez co nie jest kosztowne obliczeniowo. Zazwyczaj wykorzystuje się je w przypadku małych zbiorów danych lub gdy zasoby obliczeniowe są ograniczone. W przypadku dostrajania, usuwa się końcowe warstwy sieci, zastępując je własną nienauczoną strukturą, dopasowaną do aktualnego, nowo rozwiązywanego zadania, a następnie przeprowadza się proces douczenia z wykorzystaniem nowej bazy danych. Douczenie może dotyczyć tylko nowo dodanych warstw jak i wybranych warstw modelu bazowego przy jednoczesnym zamrożeniu (ang. *freeze*) pozostałych. Zamrożenie warstw ma na celu uniemożliwienie aktualizacji ich wag podczas uczenia [86]. Ze względu na proces douczenia dostrajanie wymaga większych zasobów obliczeniowych i dłuższego czasu uczenia w porównaniu z ekstrakcją cech. Jednocześnie technika dostrajania jest bardziej elastyczna i może potencjalnie osiągać lepsze wyniki dokładności w problemach klasyfikacyjnych [90, 91].

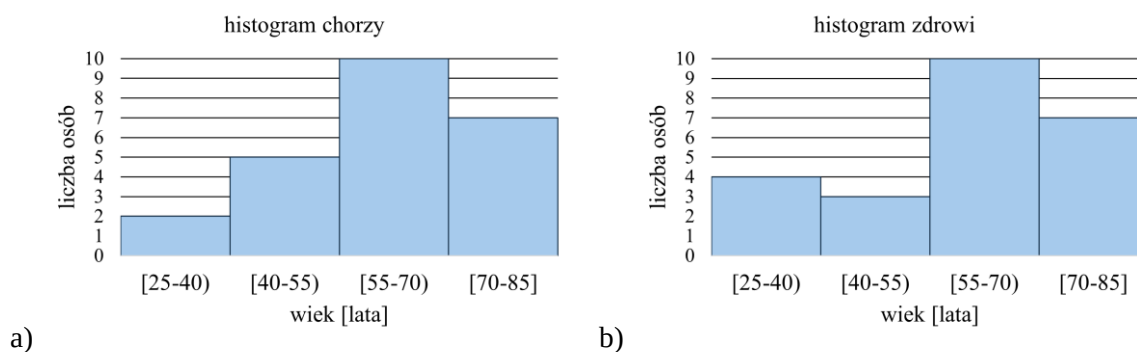
4. Rejestracja pisma odręcznego

4.1. Charakterystyka materiału badawczego

Dane obrazowe pisma odręcznego do analizy pozyskane zostały w ramach badań klinicznych przeprowadzonych przez zespół medyczny w Klinice i Katedrze Neurologii Warszawskiego Uniwersytetu Medycznego po wydaniu zgody na ich realizację przez właściwą Komisję Bioetyczną. Baza danych zawierająca uzyskane w wyniku badań dane obrazowe, będzie w niniejszej pracy oznaczana skrótem KN_19. W badaniach wzięło udział łącznie 48 osób. Wśród badanych były 24 osoby zdrowe HC (11 kobiet i 13 mężczyzn) i 24 ze stwierdzoną chorobą Parkinsona PD (16 kobiet i 8 mężczyzn). W grupie pacjentów PD ze stwierdzoną chorobą Parkinsona znajdowały się osoby w różnym stadium zaawansowania choroby, chorujące od kilku do kilkunastu lat z objawami występującymi jednostronnie (11 osób) lub obustronnie (13 osób). Co ważne, pacjenci zostali farmakologicznie przygotowani do badań w taki sposób, aby ich stan odpowiadał dolnemu zakresowi stosowanej w diagnostyce skali UPDRS, gdzie nasilenie objawów jest niewielkie przez co standardowe metody diagnostyki charakteryzują się małą czułością. Wszystkie osoby badane były praworęczne. Podstawowe charakterystyki obu grup przedstawia Tabela 4.1. Na rysunku 4.1 przedstawiono histogramy wieku uczestników badania.

Tabela 4.1. Podstawowe charakterystyki grupy badawczej (PD) i grupy kontrolnej (HC)

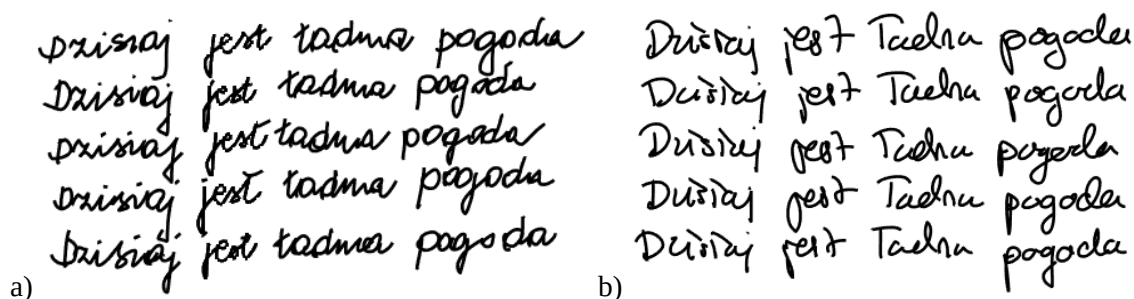
	płeć		wiek			
	kobiety	mężczyźni	średnia	odch. std.	min	max
PD	16	8	59,7	15,4	28	84
HC	11	13	58,8	16,5	26	84



Rys. 4.1. Histogramy wieku uczestników badania: a) osoby chore, b) osoby zdrowe

Każdy z pacjentów podczas standardowej wizyty lekarskiej w Klinice został poinformowany o zakresie i celu badań naukowych oraz potwierdził dobrowolność udziału w nich i podpisał świadomą zgodę na rejestrację i przetwarzanie uzyskanych danych biometrycznych. Szczegółowe informacje kliniczne i demograficzne na temat uczestników badania zamieszczono w załączniku 1.

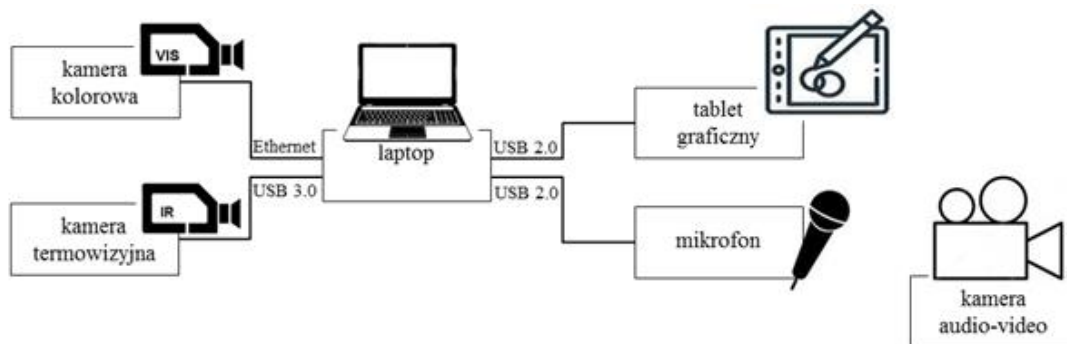
Podczas rejestracji próbek pisma za pomocą tabletu graficznego pacjenci siedzieli przy stole w wygodnej pozycji. Na tablet nałożono czystą kartkę A4, umożliwiając tym samym obserwację tworzonego pisma. W trakcie badań pacjenci zostali poproszeni o pięciokrotny zapis zdania: „Dzisiaj jest ładna pogoda”. Aby możliwe było zarejestrowanie zmian wynikających z obecności choroby i pisanie w sposób spontaniczny, wszystkie zdania były pisane bez przerwy jedno pod drugim w ramach tej samej rejestracji. Również rozmiar oraz rozmieszczenie pisma na kartce był dowolny i zależał jedynie od preferencji osoby piszącej. Przykładowe próbki pisma zarejestrowane podczas badania przedstawiono na rysunku 4.2.



Rys. 4.2. Przykładowe próbki pisma ręcznego uzyskane podczas badania: a) osoba chora, b) osoba zdrowa

4.2. Tablet graficzny jako narzędzie do rejestracji pisma odręcznego

Do rejestracji pisma wykorzystano tablet graficzny będący wraz z autorskim oprogramowaniem elementem stanowiska do pomiaru multimodalnych danych na potrzeby obiektywnej oceny choroby Parkinsona. W skład stanowiska oprócz tabletu graficznego wchodziły również: kamera podczerwieni, kamera pracująca w zakresie widzialnym, mikrofon oraz przenośny komputer. Na rysunku 4.3 zamieszczono schemat blokowy wspomnianego stanowiska.



Rys. 4.3. Schemat blokowy opracowanego stanowiska do rejestracji danych [92]

W skład rejestrowanych danych wchodził obraz wideo twarzy z zakresu światła widzialnego i podczerwieni do badania hipomimii, obrazy wideo kończyn do badania regularności ruchu stukania palcami, regularności ruchu dłoni, sprawności kończyn dolnych i chodu, próbki sygnału mowy spontanicznej i wymuszonej do oceny siły głosu, jego brzmienia i zrozumiałości oraz próbki pisma odręcznego, które zostały wykorzystane jako baza danych na potrzeby badań zawartych w poniższej rozprawie doktorskiej. Dokładny opis stanowiska wraz z charakterystyką uzyskanych za jego pośrednictwem rejestracji znajduje się w publikacji [92].

Wykorzystanym do rejestracji pisma odręcznego tabletem był, zaprezentowany na rysunku 4.4 tablet graficzny Intuos Pro Paper Edition PTH-860 firmy WACOM. Tablet umożliwia tworzenie obrazu cyfrowego na jego powierzchni za pomocą piórka dotykowego. Tablet ma wymiary $430,4 \times 287 \times 8$ mm, w tym obszar aktywny o wymiarach 311×216 mm i waży 1300 g. Za pomocą specjalnego klipsa możliwe jest przypięcie do jego powierzchni kartki papieru w formacie A4. Używane do pisania piórko to Pro Pen 2 firmy WACOM o długości 157 mm i średnicy 15 mm. Jego wymiary i waga (15 g) sprawiają, że przypomina on zwykły długopis, przez co używanie go przy obsłudze tabletu jest bardzo wygodne. Tablet umożliwia rejestrację współrzędnych lokalizacji pióra na powierzchni tabletu, a także siły nacisku z 13-bitową rozdzielczością oraz kąta nachylenia pióra w trakcie pisania. Rejestracja powyższych parametrów odbywa się z częstotliwością 200 Hz. Tablet jest zasilany z wbudowanego akumulatora litowo-jonowego, który umożliwia 12 godzinny czas pracy bez potrzeby ładowania. Urządzenie można podłączyć do komputera z wykorzystaniem interfejsu USB lub skomunikować je z komputerem poprzez bezprzewodowy interfejs Bluetooth [93].



Rys. 4.4. Tablet graficzny Intuos Pro Paper Edition firmy WACOM [94] oraz jego wykorzystanie na stanowisku do rejestracji pisma

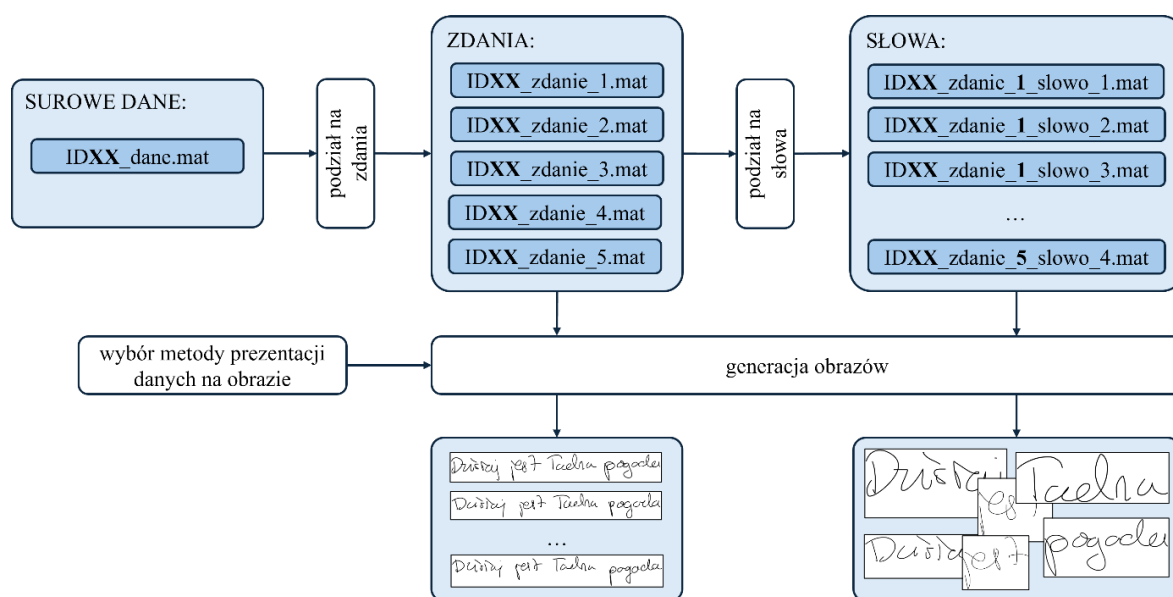
Zasada działania tabletu graficznego opiera się na wykorzystaniu zjawiska indukcji elektromagnetycznej. Powierzchnia tabletu stanowi matrycę cewek rozmieszczonych w dwóch prostokątnych osiach (X oraz Y), które pełnią funkcję zarówno nadajników, jak i odbiorników sygnałów elektromagnetycznych. Piórko dotykowe, używane do obsługi urządzenia, zawiera cewkę indukcyjną zlokalizowaną w pobliżu końcówki rysika. Po aktywacji tabletu, oscylator

generuje zmienne pole elektromagnetyczne o określonej częstotliwości, które indukuje prąd elektryczny w cewce pióra, zasilając tym samym jego układy elektroniczne i umożliwiając pomiar takich parametrów, jak siła nacisku czy kąt nachylenia rysika względem powierzchni. Precyzyjna lokalizacja położenia końcówki rysika jest realizowana na podstawie analizy sygnałów zwrotnych generowanych przez pióro i odbieranych przez cewki matrycy. Zmiany napięcia wykrywane w określonych segmentach siatki umożliwiają triangulację pozycji w osiach X i Y z bardzo wysoką dokładnością. Ponadto, pomiary są wykonywane w sposób ciągły i z wysoką częstotliwością, co pozwala na rejestrację dynamicznych ruchów pióra w czasie rzeczywistym. Przetworzone dane dotyczące pozycji, nacisku oraz kąta nachylenia są przesyłane do komputera, gdzie zostają zapisane w postaci cyfrowej, co umożliwia natychmiastowe odwzorowanie ruchów pióra na ekranie.

Szczegółowy opis autorskiego oprogramowania do rejestracji pisma z wykorzystaniem tabletu graficznego zamieszczono w załączniku 2.

5. Wstępne przetwarzanie danych

Otrzymane w wyniku rejestracji dane zapisane do pliku pod nazwą: *IDXX_dane.mat*, gdzie *XX* oznacza numer identyfikacyjny pacjenta, mają postać tabelaryczną i wymagają przetworzenia celem otrzymania na ich podstawie obrazów, które mogłyby posłużyć jako dane wejściowe dla konwolucyjnej sieci neuronowej. Dodatkowo, wykorzystanie obrazów pojedynczych zdań i słów, celem udowodnienia postawionych w rozprawie tez, wymaga podziału zarejestrowanej macierzy na osobne pliki zawierające dane dotyczące zapisu poszczególnych elementów pisma. Proces ekstrakcji pojedynczych zdań, a następnie pojedynczych słów został zautomatyzowany, poprzez opracowanie algorytmów umożliwiających samoczynny podział zarejestrowanych plików. Schemat ideowy metody przetwarzania surowych danych zamieszczono na rysunku 5.1.



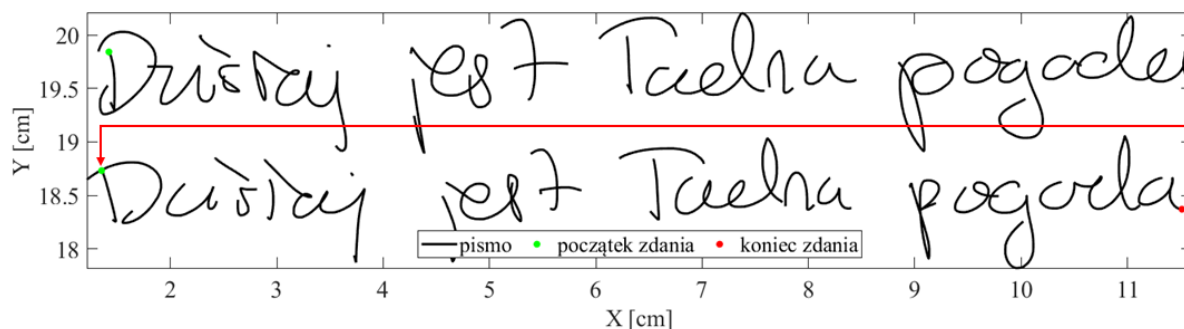
Rys. 5.1. Schemat przetwarzania danych

Do zaimplementowania opisanych w niniejszym rozdziale metod przetwarzania zarejestrowanych danych wykorzystano środowisko Matlab R2021b firmy Mathworks.

5.1. Podział całych rejestracji na poszczególne zdania

W pierwszej kolejności otrzymany w wyniku rejestracji plik należy podzielić tak, aby odseparować od siebie każde ze zdań. Dzięki temu możliwe jest odtworzenie na obrazach zapisu pojedynczych zdań. Ze względu na fakt, że zapis każdego kolejnego zdania wiąże się z przeniesieniem pióra nad powierzchnią kartki z końca jednego zdania z powrotem na

początek kartki (patrz rys. 5.2), celem podziału na pojedyncze zdania, opracowano algorytm opierający swoje działanie na zmianie wartości współrzędnej X. Dane przeszukiwane są punkt po punkcie wyszukując miejsc w których współrzędna X kolejnego punktu jest mniejsza od współrzędnej X poprzedniego punktu o pewną dobraną eksperymentalnie wartość zależną od wyznaczonej wcześniej szacunkowej szerokości całego tekstu.

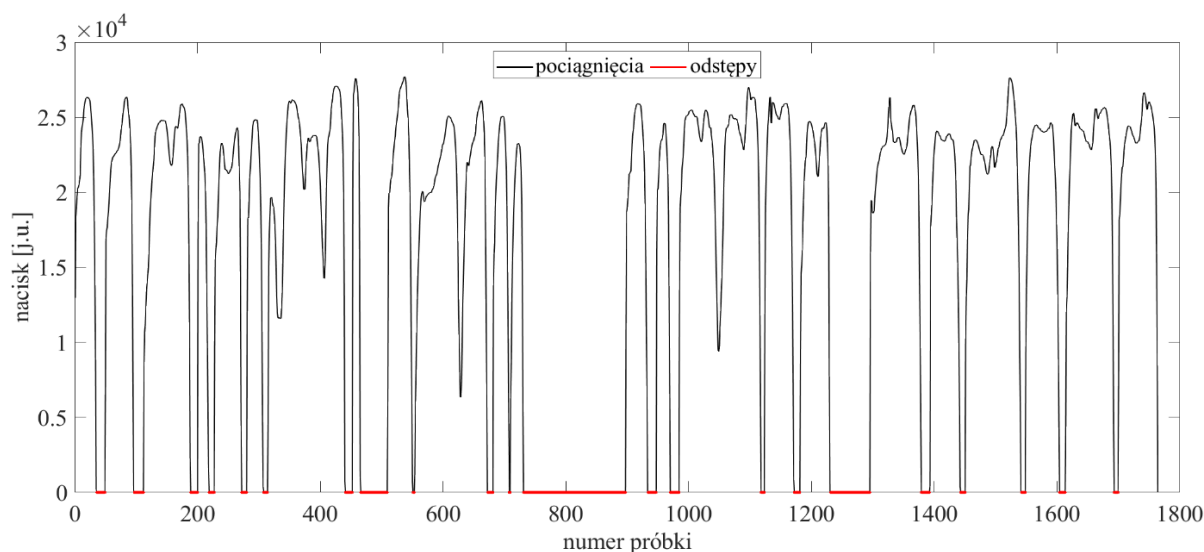


Rys. 5.2. Koncepcja działania algorytmu do podziału całej rejestracji na poszczególne zdania na przykładzie dwóch zdań

Pierwotny plik z danymi dzielony jest na pięć części na podstawie znalezionych numerów próbek stanowiących ostatnie punkty zdania poprzedniego i pierwsze punkty zdania następnego. Każdy z powstałych plików jest zapisywany pod nazwą: *IDXX_zdanie_k.mat*, gdzie *k* oznacza numer zdania.

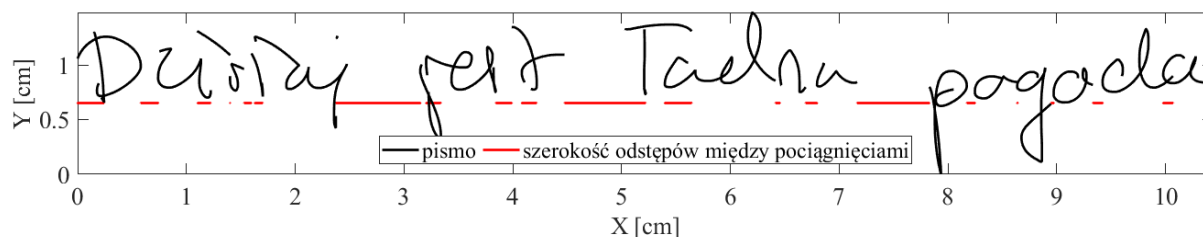
5.2. Podział zdań na słowa

W każdym zdaniu występują trzy przerwy między wyrazami, stanowiące najdłuższe odstępy między poszczególnymi pociągnięciami pióra na powierzchni tabletu. Podział zdań na słowa odbywa się na podstawie detekcji tych przerw. Odstępy i pociągnięcia zaznaczono na wykresie zamieszczonym na rysunku 5.3.



Rys. 5.3. Wykres zmian nacisku podczas zapisu jednego ze zdań z wyróżnionymi pociągnięciami i odstępami

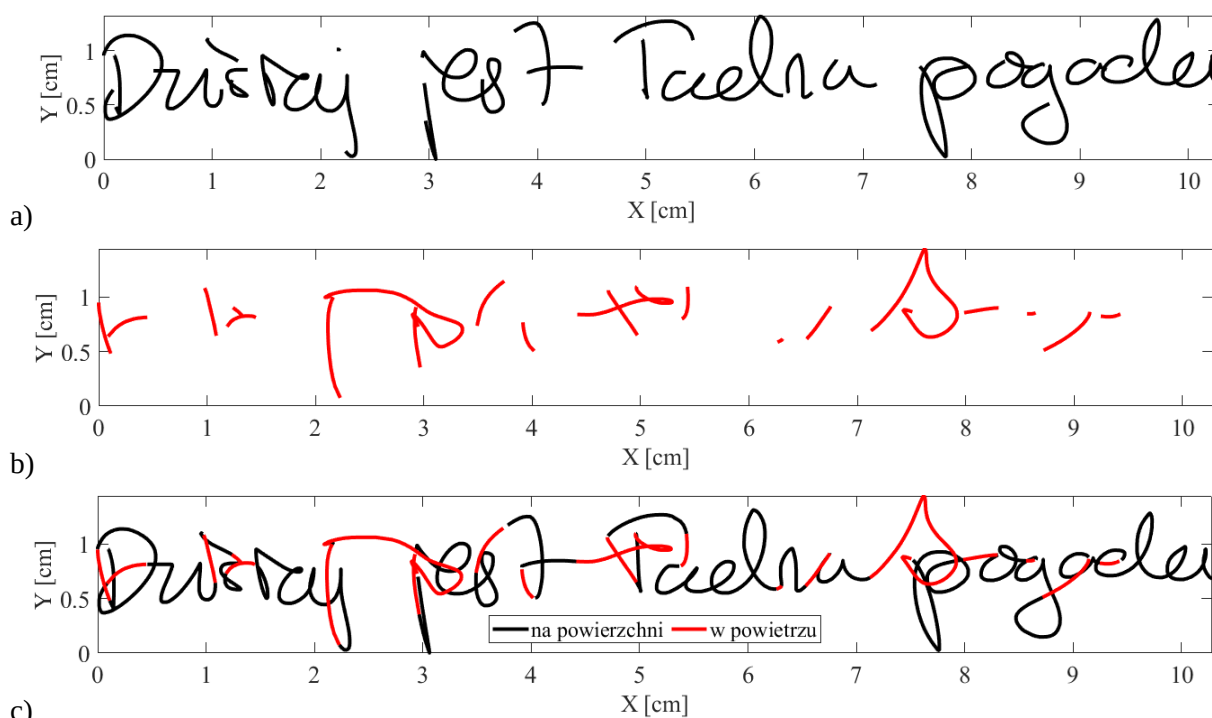
W pierwszym kroku wyznaczone zostają skrajne wartości współrzędnej X dla każdego występującego w zdaniu odstępu. Poprzez odstęp rozumie się ciąg próbek o zerowym nacisku występujący między próbkami z niezerowym naciskiem. Następnie obliczane są szerokości tych odstępów na podstawie wyznaczonych wartości współrzędnej X . Trzy najdłuższe odstępy stanowią przerwy między poszczególnymi wyrazami, na podstawie których następuje podział zdania na cztery pojedyncze słowa. Ideę podziału zdania na słowa według najdłuższych odstępów, pokazano na rysunku 5.4. Każdy z powstałych plików jest zapisywany pod nazwą: *IDXX_zdanie_k_slowo_n.mat*, gdzie k oznacza numer zdania, a n numer słowa.



Rys. 5.4. Szerokość odstępów między poszczególnymi pociągnięciami

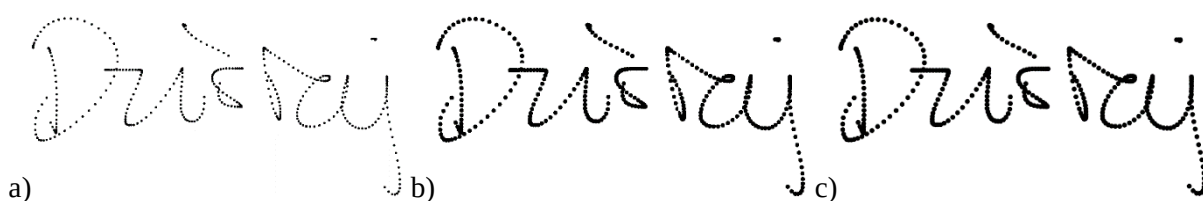
5.3. Generacja obrazów na podstawie zarejestrowanych danych

Dysponując odpowiednimi plikami zawierającymi zarejestrowane dane można odtworzyć na ekranie komputera obraz pisma zapisanego na powierzchni tabletu. W pierwszej kolejności zostają usunięte tzw. puste punkty, czyli próbki, w których nie zostały zarejestrowane żadne dane. Zjawisko wystąpienia takich punktów wynika z faktu, że urządzenie rejestrujące rozpoczęło zbieranie danych, zanim użytkownik faktycznie przystąpił do pisania (umieścił rysik na powierzchni tabletu) lub kiedy pacjent zakończył pisanie, a tablet nadal jest aktywny i kontynuuje rejestrację. Następnie, na podstawie wartości nacisku, usuwane są wszystkie punkty zebrane w momencie gdy pióro znajdowało się nad powierzchnią tabletu (patrz rysunek 5.5b). Jest to realizowane poprzez znalezienie w macierzy danych tych próbek, dla których wartość nacisku jest równa zero i ich usunięcie. Dzięki temu na obrazie zostanie odtworzony tylko ślad, który w rzeczywistości został narysowany na kartce (patrz rys. 5.5a). Bez wykonania tego kroku na obrazie byłoby widoczne zarówno rzeczywiste pismo powstałe na kartce, jak i wszystkie ruchy pióra wykonane nad powierzchnią tabletu (patrz rys. 5.5c). Ma to szczególne znaczenie podczas rejestracji procesu pisania, gdzie wymagane jest oderwanie pióra od kartki między poszczególnymi znakami.



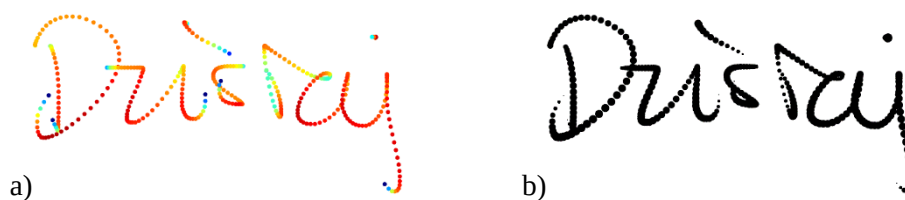
Rys. 5.5. Zapis zarejestrowanego zdania: a) tylko próbki na powierzchni tabletu, b) tylko próbki nad powierzchnią tabletu, c) wszystkie próbki

W najprostszej wersji gotowy obraz jest tworzony z wykorzystaniem wbudowanej funkcji `plot()` jako wykres zależności wartości współrzędnych X i wartości współrzędnych Y za pomocą znaczników o odpowiednim rozmiarze. Rozmiar znacznika definiowany jest poprzez parametr `'MarkerSize'`, wyrażany w punktach, gdzie 1 punkt odpowiada wartości 1/72 cala. Ostatecznie uzyskuje się obraz przedstawiający czarną grafikę na białym tle. Odtworzony obraz pojedynczego słowa z zastosowaniem różnych grubości punktów zaprezentowano na rysunku 5.6.



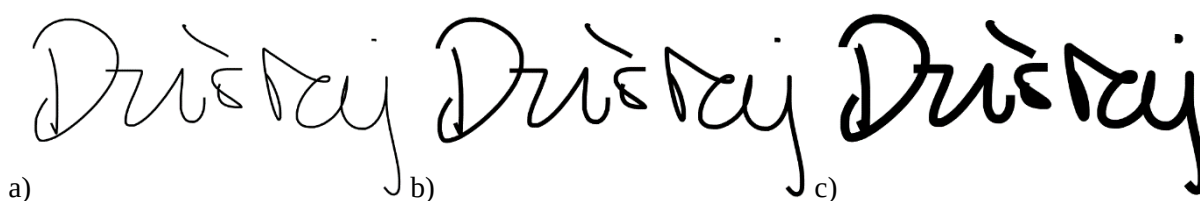
Rys. 5.6. Odtworzony obraz słowa „Dzisiaj” z zastosowaniem rozmiaru znacznika: a) 5 punktów, b) 10 punktów, c) 15 punktów

Takie podejście umożliwia zaprezentowanie w postaci obrazu również innych parametrów, jak na przykład wartości nacisku przyłożonego do powierzchni tabletu podczas pisania z wykorzystaniem odpowiedniej skali kolorystycznej lub zmieniając rozmiar znacznika w zależności od wartości nacisku. Przykładowe obrazy pojedynczego słowa z uwzględnieniem dodatkowych informacji zaprezentowano na rysunku 5.7.



Rys. 5.7. Odtworzony obraz słowa „Dzisiaj” z uwzględnieniem wartości nacisku poprzez zastosowanie: a) skali kolorystycznej, b) zmiany rozmiaru znacznika

Odtworzenie na obrazie pełnego śladu odpowiadającego rzeczywistemu zapisowi widniejącemu na kartce papieru, wymaga zastosowania dodatkowo interpolacji między poszczególnymi punktami. Aby na obrazie nie uwidocznili odcinków związanych z przenoszeniem pióra między poszczególnymi pociągnięciami, interpolację wykonuje się w ramach poszczególnych pociągnięć a nie całego zarejestrowanego zapisu. Podobnie jak w przypadku rozmiaru znacznika, grubość linii jest definiowana jako parametr *'LineWidth'* funkcji *plot()*. Odtworzony obraz pojedynczego słowa z zastosowaniem różnych grubości linii zaprezentowano na rysunku 5.8.



Rys. 5.8. Odtworzony obraz słowa „Dzisiaj” z zastosowaniem grubości linii: a) 2 punkty, b) 4 punkty, c) 8 punktów

6. Organizacja badań eksperymentalnych

Zakres wykonywanych badań, warunkowany weryfikacją postawionych hipotez, obejmuje rozpoznawanie choroby Parkinsona na podstawie obrazów prezentujących pismo odręczne w postaci: a) pojedynczych zdań; b) pojedynczych zdań pochodzących z zapisu ich sekwencji; c) pojedynczych słów pochodzących z podziału zdań na słowa.

Podczas badań i analizy wyników wykorzystywano komputer z 64-bitowym systemem Windows 10, procesorem Intel Core i7-8750H 2,20 GHz i pamięcią RAM 16 GB. Proces uczenia i testowania sieci oraz wszystkie obliczenia przeprowadzono w środowisku Matlab R2018b firmy MathWorks, z wykorzystaniem karty graficznej NVIDIA GeForce GTX 1050 4 GB. Środowisko Matlab posiada w swoich bibliotekach wiele modeli sieci konwolucyjnych, m.in. AlexNet, GoogLeNet, VGG16, ResNet18, InceptionV3, MobileNet przeszkolonych na zbiorze danych ImageNet, zawierającym obrazy należące do 1000 klas. Dzięki temu możliwe jest zaimplementowanie gotowej struktury sieci do rozwiązywania nowego zadania poprzez jej douczenie (zgodnie z zasadami przedstawionymi w podrozdziale 3.6). Zastosowana w badaniach opisanych w niniejszej rozprawie technika uczenia to stochastyczna metoda największego spadku z momentem, z wykorzystaniem zbiorów *mini batches*, gdzie w każdej iteracji zamiast pełnego gradientu funkcji celu stosuje się jego aproksymację z użyciem małego zbioru danych uczących wybranych losowo ze zbioru pełnego.

6.1. Walidacja krzyżowa

Przystępując do rozwiązania problemu klasyfikacji, dostępny zbiór danych należy podzielić na podzbiór uczący, na którym dobierane są optymalne wartości parametrów modelu i podzbiór testujący służący sprawdzeniu zdolności generalizacyjnych wytrenowanego układu. Przy podziale zestawu danych ważne jest, aby dane uczące i testujące stanowiły dwa rozłączne zbiory. Zazwyczaj zbiór dostępnych danych dzieli się w sposób losowy, z czym wiąże się problem losowości uzyskanych wyników testowania. W związku z tym, aby wiarygodnie ocenić proces klasyfikacji zalecane jest wielokrotne powtórzenie procesu uczenia i testowania z wykorzystaniem innych zbiorów uczących i testujących. Istnieją różne metody praktycznej implementacji tego podejścia. Jedną z najczęściej stosowanych jest metoda k -krotnej walidacji krzyżowej (ang. *k-fold cross validation*). Metoda ta polega na losowym podziale całego zestawu danych na k podzbiorów w przybliżeniu równej wielkości, a następnie użyciu pojedynczego zbioru jako danych testujących, a pozostałych podzbiorów jako danych uczących. Następnie proces powtarza się k razy zmieniając za każdym razem zestawy testujące i uczące. Wyniki

klasyfikacji uzyskane za pomocą poszczególnych podzbiorów są następnie uśredniane [95]. Ideę 8-krotnej walidacji krzyżowej zaprezentowano na rysunku 6.1.

	ZBIÓR DANYCH							
ITERACJA 1	TEST	TRENING	TRENING	TRENING	TRENING	TRENING	TRENING	TRENING
ITERACJA 2	TRENING	TEST	TRENING	TRENING	TRENING	TRENING	TRENING	TRENING
ITERACJA 3	TRENING	TRENING	TEST	TRENING	TRENING	TRENING	TRENING	TRENING
ITERACJA 4	TRENING	TRENING	TRENING	TEST	TRENING	TRENING	TRENING	TRENING
ITERACJA 5	TRENING	TRENING	TRENING	TRENING	TEST	TRENING	TRENING	TRENING
ITERACJA 6	TRENING	TRENING	TRENING	TRENING	TRENING	TEST	TRENING	TRENING
ITERACJA 7	TRENING	TRENING	TRENING	TRENING	TRENING	TRENING	TEST	TRENING
ITERACJA 8	TRENING	TRENING	TRENING	TRENING	TRENING	TRENING	TRENING	TEST

Rys. 6.1. Przykład 8-krotnej walidacji krzyżowej

W celu uzyskania lepszego rezultatu zastosowania tej metody zbiór danych powinien być podzielony w taki sposób, aby stosunek reprezentantów poszczególnych klas w zbiorze testującym był zbliżony do tego w zbiorze uczącym. Dobór parametru k zależy od wielkości zbioru danych i ich rodzaju. Dla małych zbiorów danych stosuje się najczęściej k z zakresu od 5 do 10, ale nie ma specjalnych wymagań co do wyboru tej wartości [96, 97]. W opisywanych w ramach niniejszej pracy badaniach, zastosowano 8-krotną walidację krzyżową. Wybór $k = 8$ został podyktowany równomiernym podziałem całego zestawu danych. Dysponując danymi pochodzącymi łącznie od 48 osób, w celu wykorzystania zbiorów testowych i uczących niezależnych od osoby, pojedynczy podzbiór testujący w walidacji krzyżowej zawierał obrazy pisma pochodzącego od 6 osób, odpowiednio od 3 osób z każdej z klas. Tym sposobem, w każdej z k iteracji do uczenia zostały wykorzystane dane pochodzące od 42 osób, a do testowania dane pochodzące od 6 osób.

6.2. Miary jakości klasyfikacji

Ocena wydajności modelu klasyfikacyjnego opiera się na zgodności estymowanej etykiety klasy z etykietą rzeczywistą. Najczęściej wykorzystywanym sposobem prezentacji wyników klasyfikacji jest zastosowanie macierzy błędów (ang. *confusion matrix*), w której elementy kolumn reprezentują liczbę wzorców należących do kolejnych klas, a elementy wierszy – liczbę wzorców rozpoznanych przez klasyfikator jako dana klasa. W przypadku klasyfikacji binarnej, jakim jest rozpoznawanie osób chorych i osób zdrowych, macierz błędów ma postać, jak na rysunku 6.2.

		STAN FIZYCZNY	
		CHORZY	ZDROWI
WYNIK KLASYFIKACJI	CHORZY	TP – ang. true positive	FP – ang. false positive
	ZDROWI	FN – ang. false negative	TN – ang. true negative

Rys. 6.2. Macierz błędów dla przypadku rozpoznawania klas: chory i zdrowy

Elementy diagonalne tej macierzy oznaczają liczbę osób TP prawidłowo sklasyfikowanych jako chore i liczbę osób TN poprawnie sklasyfikowanych jako zdrowe. Komórki pozadiagonalne zawierają liczbę osób sklasyfikowanych niepoprawnie. FP oznacza zdrowe przypadki błędnie sklasyfikowane jako chore, natomiast FN osoby chore błędnie przypisane do grupy osób zdrowych [98]. Na podstawie macierzy błędów definiuje się szereg wskaźników jakości klasyfikacji, dzięki którym możliwe jest porównanie wydajności różnych modeli klasyfikacyjnych. Do najczęściej używanych miar jakości klasyfikacji w kontekście diagnozowania choroby należą:

Dokładność (ang. *accuracy*) to parametr ogólnej jakości klasyfikacji, reprezentujący liczbę poprawnie sklasyfikowanych przypadków w odniesieniu do wszystkich przypadków. Dokładność wyrażana jest wzorem 6.1:

$$\text{Dokładność} = \frac{TP + TN}{TP + TN + FP + FN} \quad (6.1)$$

Czułość (ang. *recall, sensitivity*) jest to stosunek liczby poprawnie rozpoznanych przypadków chorych do liczby wszystkich przypadków chorych. Czułość można przedstawić wzorem 6.2:

$$\text{Czułość} = \frac{TP}{TP + FN} \quad (6.2)$$

Specyficzność (ang. *specificity*) definiowana jest jako stosunek liczby poprawnie rozpoznanych przypadków zdrowych do liczby wszystkich przypadków zdrowych i wyrażana za pomocą wzoru 6.3:

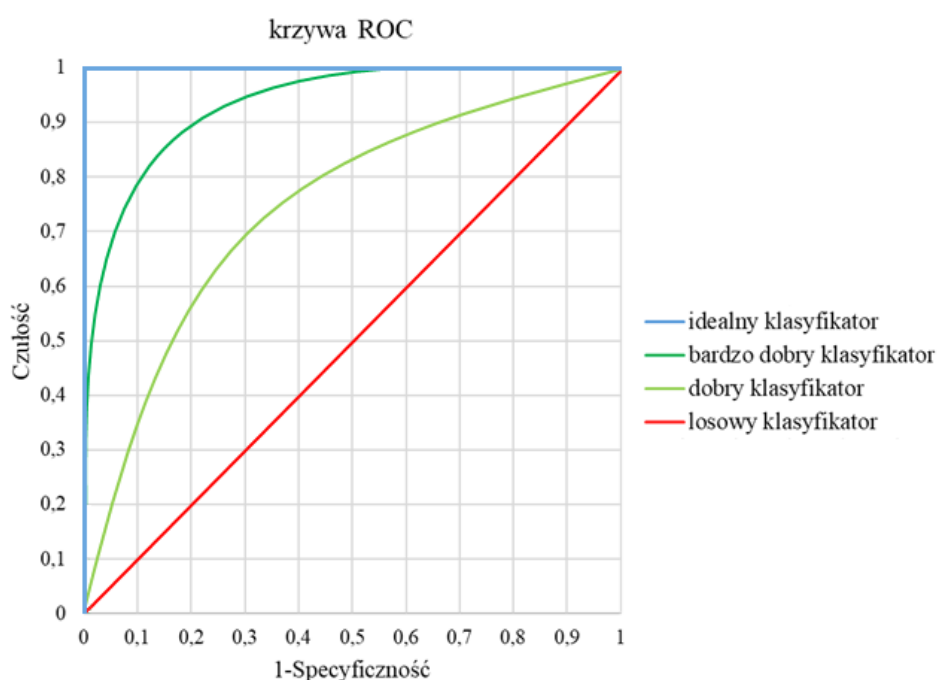
$$\text{Specyficzność} = \frac{TN}{TN + FP} \quad (6.3)$$

Precyzja (ang. *precision*) określa jaka część przypadków zdiagnozowanych jako chore jest rzeczywiście chora. Precyzję można wyrazić za pomocą wzoru 6.4:

$$\text{Precyzja} = \frac{TP}{TP + FP} \quad (6.4)$$

Warto zauważyć, że wartości powyższych wskaźników będą zależały od przyjętego progu oceny danego przypadku. Możliwa jest modyfikacja wartości wskaźników poprzez zmianę

wartości progu decydującego o przyporządkowaniu do danej klasy, jednak maksymalizacja jednego ze wskaźników jest często sprzężona z pogorszeniem innego. Wpływ zmian progu decyzyjnego na wynik działania klasyfikatora można zobrazować za pomocą krzywej ROC (ang. *Receiver Operating Characteristics*). Krzywa ROC prezentuje relację między czułością a specyficznością dla różnych wartości progu. Na osi X krzywej ROC wyznacza się tzw. prawdopodobieństwo fałszywego alarmu ($1 - \text{Specyficzność}$), a na osi Y - Czułość. Przebieg krzywej ROC świadczy o jakości danego klasyfikatora. Dobry klasyfikator powinien mieć krzywą ROC położoną tak blisko, jak to możliwe lewego górnego punktu o współrzędnych (0,1). Im dalej krzywa ROC przebiega od tego punktu tym gorszy jest klasyfikator (patrz rys. 6.3). Prosta łącząca punkty (0,0) i (1,1), będzie charakteryzować klasyfikator losowy, w którym wynik klasyfikacji odbywa się na drodze losowania [99].



Rys. 6.3. Prezentacja oceny jakości klasyfikatora na podstawie krzywej ROC

Przesuwanie się wzdłuż krzywej ROC od lewego dolnego do prawego górnego rogu powoduje wzrost czułości modelu kosztem pogorszenia jego specyficzności. Optymalny punkt odcięcia (ang. *optimal cut-off*) będzie taką wartością progową, która zapewni kompromis pomiędzy czułością i specyficznością. Istnieje kilka różnych metod doboru optymalnej wartości punktu odcięcia. Najpopularniejszym jest kryterium punktu będącego najbliższym punktu (0,1), gdzie jako punkt odcięcia wybiera się taką wartość progową x , która minimalizuje funkcję, przedstawioną zależnością 6.5 [100]:

$$R(x) = \sqrt{(1 - \text{Czułość}(x))^2 + (1 - \text{Specyficzność}(x))^2} \quad (6.5)$$

Innym kryterium jest indeks Youdena (ang. the Youden index), gdzie jako punkt odcięcia wybiera się taką wartość progową x , która maksymalizuje funkcję, przedstawioną zależnością 6.6 [101]:

$$J(x) = \text{Czułość}(x) - (1 - \text{Specyficzność}(x)) \quad (6.6)$$

Jest to równoważne ze znalezieniem punktu krzywej ROC o największej odległości w pionie od prostej reprezentującej klasyfikator losowy. Możliwe jest również zastosowanie metody prawdopodobieństwa zgodności (ang. *concordance probability method*), gdzie za optymalny punkt odcięcia uznaje się ten, który wyznacza prostokąt o największym polu. Wybór wartości progowej x odbywa się poprzez maksymalizację funkcji, przedstawionej zależnością 6.7 [100]:

$$C(x) = \text{Czułość}(x) * \text{Specyficzność}(x) \quad (6.7)$$

7. Rozpoznawanie choroby Parkinsona na podstawie pojedynczego zdania

W niniejszym rozdziale przedstawiona została klasyfikacja binarna osób chorych na chorobę Parkinsona i osób zdrowych na podstawie pisma odręcznego przedstawiającego pojedyncze zdanie. Celem badania była weryfikacja wyników dokładności rozpoznania choroby na próbkach pisma pochodzących z własnej bazy danych i zestawienie ich z wynikami analogicznych badań opisanych w literaturze, przeprowadzonych z wykorzystaniem danych pochodzących z bazy PaHaW. Jednym z zadań piśmienniczych dostępnym w tym zbiorze danych jest pojedyncze zdanie w języku czeskim: „*Tramwaj dnes už nepojede.*” (z czes. *Tramwaj dzisiaj nie pojeździe.*). Wyniki rozpoznawania choroby Parkinsona na podstawie tego zdania mogą stanowić porównanie dla uzyskanych w ramach niniejszej rozprawy wyników. W tym rozdziale wykorzystane zostaną dwa podejścia do techniki *transfer learningu*, scharakteryzowane w podrozdziale 3.6: ekstrakcja cech i dostrajanie z użyciem sieci AlexNet.

7.1. Metodyka badań

7.1.1. Struktura sieci AlexNet

Struktura sieci AlexNet składa się z pięciu warstw konwolucyjnych i trzech warstw pełnego połączenia. W pierwszej, drugiej i piątej warstwie wykonywana jest dodatkowo operacja *poolingu* typu *max pooling*. W sieci AlexNet jako jednej z pierwszych sieci konwolucyjnych, wykorzystano funkcje aktywacji ReLU zamiast funkcji sigmoidalnej, które były standardem dla poprzednich sieci neuronowych. Pierwsza warstwa w pełni połączona *fc6* współpracująca z operacją ReLU jest warstwą zawierającą 4096 neuronów. Druga warstwa w pełni połączona *fc7* jest warstwą ukrytą klasyfikatora końcowego zasilającą warstwę wyjściową *fc8* składającą się z 1000 neuronów, pozwalającą na rozpoznanie 1000 klas. Zastosowany w strukturze klasyfikator jest zakończony warstwą *softmax*. Sieć AlexNet została wytrenowana z użyciem bazy danych ImageNet składającej się z 1,2 mln obrazów należących do 1000 różnych klas. Wymiar wejściowy obrazów sieci AlexNet jest równy $227 \times 227 \times 3$. W pierwszej warstwie konwolucyjnej (*conv1*) przeprowadzona jest operacja konwolucji z użyciem 96 filtrów o wymiarze $11 \times 11 \times 3$ z krokiem *stride* równym 4. Druga warstwa (*conv2*) otrzymuje na wejściu znormalizowany i połączony sygnał wyjściowy pierwszej warstwy i przeprowadza filtrację za pomocą 256 filtrów o wymiarze $5 \times 5 \times 48$. Trzecia warstwa (*conv3*) wykorzystuje 384 filtry o rozmiarze $3 \times 3 \times 256$ do przeprowadzenia operacji konwolucji na sygnałach wyjściowych drugiej warstwy. Czwarta warstwa konwolucyjna (*conv4*) zawiera 384 filtry

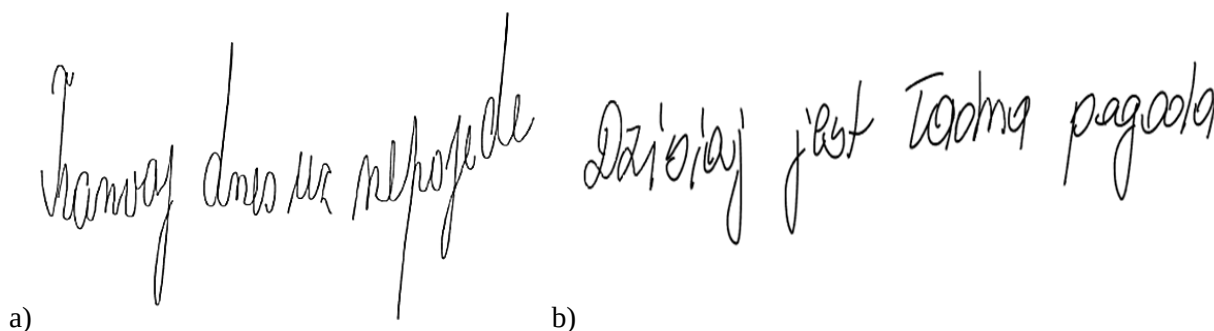
o wymiarach $3 \times 3 \times 192$, a piąta (*conv5*) 256 filtrów o wymiarach $3 \times 3 \times 192$. Na wyjściu piątej warstwy po operacji pooling'u otrzymuje się 256 map cech o wymiarze 4×4 , które następnie w wyniku konkatencji są przekształcone w wektor cech składający się z 4096 elementów, podanych na poszczególne neurony pierwszej warstwy w pełni połączonej *fc6* [102]. Strukturę sieci wraz z opisem poszczególnych warstw zamieszczono w tabeli 7.1.

Tabela 7.1. Struktura sieci AlexNet

warstwa	opis
<i>data</i>	dane wejściowe o wymiarze 227x227x3
<i>conv1</i>	konwolucja, 96 filtrów o wymiarze 11x11x3, krok 4
<i>pool1</i>	max pooling, maska 3x3, krok 2
<i>conv2</i>	konwolucja, 256 filtrów o wymiarze 5x5x48, krok 1, zero padding 2
<i>pool2</i>	max pooling, maska 3x3, krok 2
<i>conv3</i>	konwolucja, 384 filtry o wymiarze 3x3x256, krok 1, zero padding 1
<i>conv4</i>	konwolucja, 384 filtry o wymiarze 3x3x192, krok 1, zero padding 1
<i>conv5</i>	konwolucja, 256 filtrów o wymiarze 3x3x192, krok 1, zero padding 1
<i>pool5</i>	max pooling, maska 3x3, krok 2
<i>fc6</i>	warstwa pełnego połączenia, 4096 neuronów
<i>fc7</i>	warstwa pełnego połączenia, 4096 neuronów
<i>fc8</i>	warstwa pełnego połączenia, 1000 neuronów
<i>prob</i>	softmax
<i>output</i>	klasyfikacja, 1000 klas

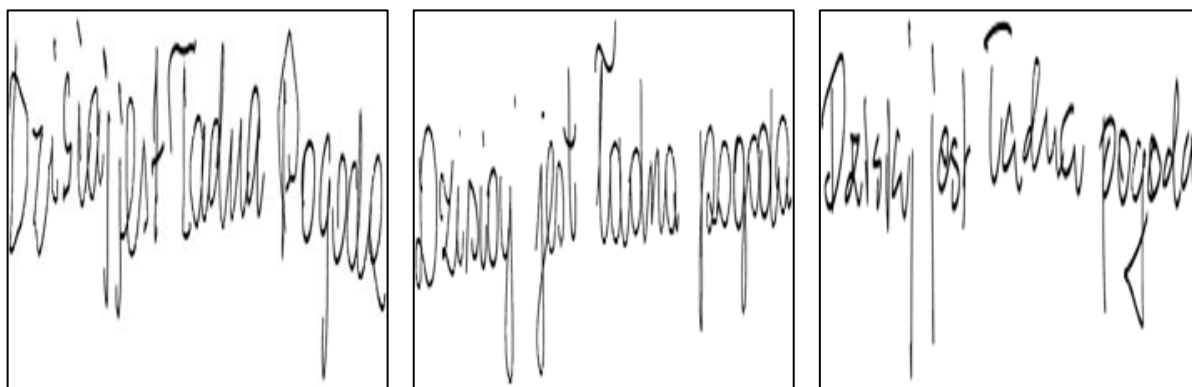
7.1.2. Przygotowanie danych

W pierwszej kolejności przeprowadzony został eksperyment analogiczny do opisanego w [79]. W tej publikacji wyodrębniono wektor cech z warstwy w pełni połączonej *fc7* wstępnie wyszkolonej sieci AlexNet na zbiorze danych ImageNet. Otrzymany wektor cech podawano następnie na klasyfikator typu SVM. Nie zastosowano tutaj żadnych metod augmentacji. W pracy znajdują się między innymi wyniki otrzymane dla surowych obrazów zawierających pełny ślad pisma, odzwierciedlający jego wygląd na kartce papieru, czyli z zastosowaniem interpolacji. We wspomnianej publikacji brakuje jednak informacji na temat grubości linii zastosowanej przy generacji obrazów. W związku z tym podczas generacji obrazów zdań zasugerowano się kolejną pracą [81], w której przeprowadzono podobne do [79] badania i zamieszczonym w niej przykładowym obrazem zdania. Na tej podstawie zdecydowano się zastosować linię o grubości 1 pkt. Na rysunku 7.1 zamieszczono porównanie obrazów wykorzystanych w artykule [81] i w niniejszej pracy.



Rys. 7.1. Zestawienie przykładowych obrazów prezentujących pojedyncze zdanie: a) z bazy danych PaHaW w artykule [81], b) z autorskiej bazy danych KN_19

W tej części badań wykorzystano pojedyncze zdania zapisane przez każdą z osób badanych. Wykreślone obrazy prezentujące pojedyncze zdania, przewymiarowano tak, aby ich rozmiar odpowiadał wymaganemu przez sieć AlexNet – $227 \times 227 \times 3$. Przykładowe obrazy zaprezentowano na rysunku 7.2.



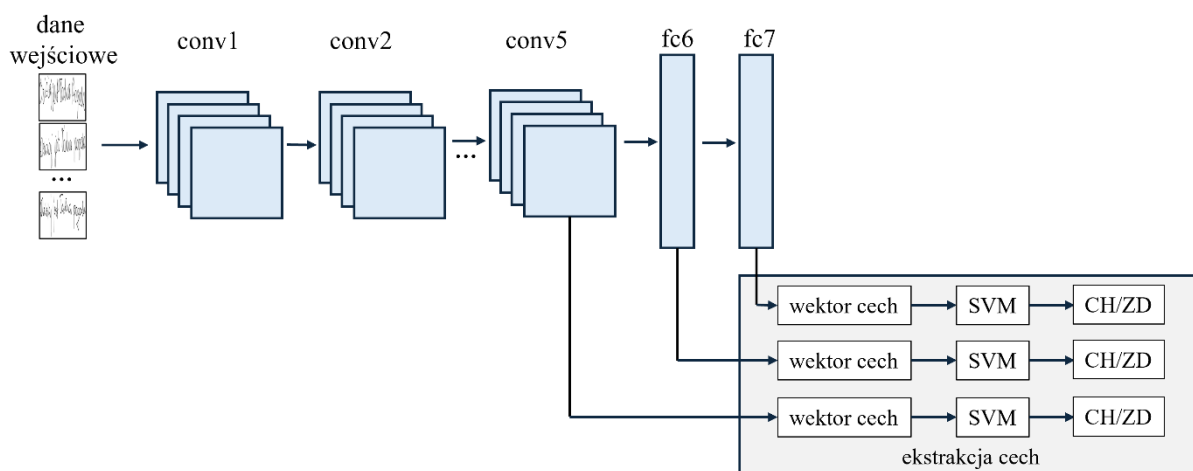
Rys. 7.2. Przykładowe obrazy pojedynczych zdań użyte jako dane wejściowe sieci

Następnie, zbiór 48 obrazów został losowo podzielony na 8 równych podzbiorów, z których każdy zawierał 3 obrazy pochodzące od osób chorych i 3 pochodzące od osób zdrowych. W każdej z 8 iteracji, 42 obrazy zostały użyte do uczenia sieci, a 6 do testowania.

7.1.3. Ekstrakcja cech

W podejściu ekstrakcji cech, wektor cech jest pobierany jako dane wyjściowe jednej z warstw, najczęściej w pełni połączonej, a wszystkie kolejne warstwy zostają usunięte. Otrzymany wektor cech jest przekazywany do innego klasyfikatora. W publikacji [79] wektor cech został wyodrębniony z przedostatniej warstwy pełnego połączenia *fc7*. W ramach niniejszej pracy przeprowadzono również ekstrakcję cech z warstw *fc6* oraz ostatniej warstwy konwolucyjnej *conv5*. W publikacji [81], na podstawie wektora cech pochodzącego właśnie z warstwy *conv5* uzyskano wysokie wyniki rozpoznania. Do przeprowadzenia klasyfikacji na podstawie otrzymanego wektora cech użyto najczęściej wykorzystywany w tym celu klasyfikator SVM. W publikacji [79] brakuje szczegółowych informacji na temat zastosowanej funkcji jądra i wartości parametrów klasyfikatora SVM, dlatego zdecydowano się użyć

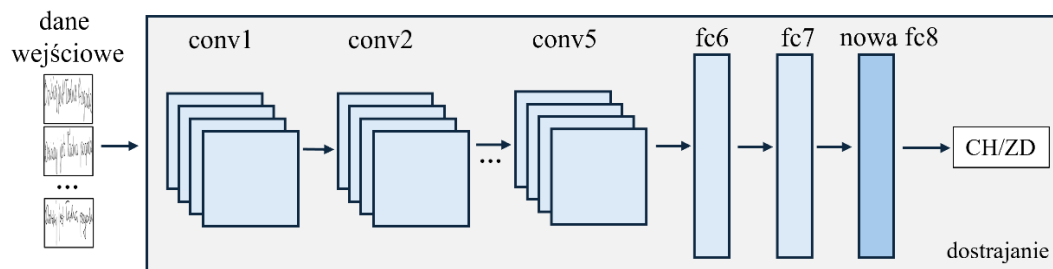
powszechnie stosowanej w tego typu zadaniach, radialnej funkcji jądra o wartości parametru $\gamma = 1/n$, gdzie n oznacza liczbę cech oraz wartości stałej regularyzacji $C = 1$. Ekstrakcję wektora cech oraz klasyfikację osób chorych i zdrowych na jego podstawie przeprowadzono w dwójnasób: z wykorzystaniem sieci AlexNet, przetrenowanej tylko na zbiorze danych ImageNet oraz z wykorzystaniem tej sieci, ale douczanej autorskim zbiorem danych KN_19 w procesie dostrajania. Schemat wykonywania eksperymentu opartego na ekstrakcji cech zaprezentowano na rysunku 7.3.



Rys. 7.3. Schemat wykonywania eksperymentu opartego na ekstrakcji cech

7.1.4. Dostrajanie

W przypadku dostrajania pobraną strukturę sieci AlexNet, należy wstępnie zmodyfikować, celem dopasowania jej do zadania rozpoznawania choroby Parkinsona. Oryginalna sieć została zaprojektowana do klasyfikacji obiektów należących do 1000 klas. Celem przystosowania sieci do zadania rozpoznawania choroby Parkinsona, usunięto jej ostatnie warstwy tj. warstwę w pełni połączoną *fc8* oraz *softmax* i zastąpiono je nowymi, które wraz z pozostałymi warstwami zostały douczone w procesie uczenia sieci. Nowa warstwa w pełni połączona zawiera dwa neurony, tak, aby wyróżniała dwie kategorie: „CHORY” i „ZDROWY”. Schemat wykonywania eksperymentu z wykorzystaniem dostrajania zaprezentowano na rysunku 7.4.



Rys. 7.4. Schemat wykonywania eksperymentu opartego na dostrajaniu

7.1.5. Uczenie sieci

Zmodyfikowaną strukturę, wstępnie przetrenowanej sieci AlexNet, douczono autorskim zbiorem danych, zachowując przyjęty schemat 8-krotnej walidacji krzyżowej. Ze względu na różny charakter zadania źródłowego bazy ImageNet i zadania rozpoznawania choroby Parkinsona na podstawie próbek pisma, w trakcie uczenia sieci, żadna z warstw nie była zamrożona. Oznacza to, że wszystkie warstwy podlegały procesowi aktualizacji wag. Nie mniej jednak wzmocniono uczenie się ostatniej warstwy, dodanej na etapie dopasowywania struktury sieci do zadania, poprzez dwukrotne zwiększenie domyślnej wartości współczynnika szybkości uczenia się wag *'WeightLearnRateFactor'* w tej warstwie. Wobec niewielkiej liczby danych uczących rozmiar podzbioru uczącego *mini batch* jest równy całkowitej liczbie przypadków. Jako warunek końca uczenia się sieci przyjęto osiągnięcie dokładności rozpoznania na danych uczących wynoszące 100 % dla pięciu kolejnych iteracji uczących. Pozostałe wartości parametrów uczenia sieci zaprezentowano w tabeli 7.2.

Tabela 7.2. Zastosowane wartości parametrów uczenia sieci

nazwa parametru (MATLAB)	definicja parametru	wartość
<i>Momentum</i>	współczynnik momentu	0,9
<i>InitialLearnRate</i>	początkowy współczynnik uczenia	0,0001
<i>MiniBatchSize</i>	rozmiar podzbioru uczącego <i>mini batch</i>	42
<i>MaxEpochSize</i>	maksymalna liczba epok uczących	100

7.2. Wyniki eksperymentu

7.2.1. Rozpoznawanie na podstawie pierwszego zdania

W tej części badań, jako dane dla sieci, wykorzystano tylko pierwsze zdania zapisane przez każdą z osób badanych. Proces zapisu pierwszego zdania z całej sekwencji pięciu zdań jest adekwatny do zapisu tylko jednego zdania, jak miało to miejsce w przypadku bazy danych PaHaW. W tabeli 7.3 zaprezentowano wyniki rozpoznania choroby Parkinsona na podstawie próbek pisma ze zbioru testującego, w postaci średniej dokładności z 8 iteracji walidacji krzyżowej oraz odchylenia standardowego.

Tabela 7.3. Wyniki rozpoznania choroby Parkinsona na podstawie pierwszego zdania

ekstrakcja cech – AlexNet - ImageNet			
	<i>conv5</i>	<i>fc6</i>	<i>fc7</i>
Dokładność	0,50 ± 0,00	0,50 ± 0,00	0,52 ± 0,21
Precyzja	-	-	-
Czułość	0,00 ± 0,00	0,88 ± 0,35	0,46 ± 0,35
Specyficzność	1,0 ± 0,0	0,12 ± 0,35	0,58 ± 0,30

Tabela 7.3. Wyniki rozpoznania choroby Parkinsona na podstawie pierwszego zdania (cd.)

ekstrakcja cech – AlexNet – ImageNet + KN_19			
	<i>conv5</i>	<i>fc6</i>	<i>fc7</i>
Dokładność	0,50 ± 0,00	0,479 ± 0,059	0,60 ± 0,15
Precyzja	-	0,488 ± 0,035	0,62 ± 0,20
Czułość	0,25 ± 0,46	0,88 ± 0,17	0,67 ± 0,31
Specyficzność	0,75 ± 0,46	0,08 ± 0,15	0,54 ± 0,25
dostrajanie			
Dokładność	0,60 ± 0,15		
Precyzja	0,58 ± 0,28		
Czułość	0,62 ± 0,38		
Specyficzność	0,58 ± 0,24		

W przypadku ekstrakcji cech z użyciem sieci tylko wstępnie przeszkolonej na bazie danych ImageNet, wyniki rozpoznania są na poziomie klasyfikatora losowego. Tylko w przypadku ekstrakcji z warstwy w pełni połączonej *fc7* uzyskano wynik dokładności rozpoznania powyżej 50 %. Wyniki te są zbliżone do uzyskanych w [79], gdzie wyniki dokładności rozpoznania dla analogicznego eksperymentu, zależnie od użytych reprezentacji wahał się w granicach 48 – 51 %. (patrz rozdział 2.4)

Ekstrakcja cech na podstawie modelu doszkołonego nową bazą danych KN_19, w przypadku wektora cech wyodrębnionego z warstw *conv5* i *fc6* nie przyniosła oczekiwanych rezultatów. Natomiast najwyższy wynik dokładności rozpoznania wynoszący 60 % uzyskano w przypadku dostrajania oraz w przypadku ekstrakcji cech z doszkołonego modelu z warstwy *fc7*. Na rysunku 7.5 zaprezentowano macierze błędów uzyskane na podstawie ekstrakcji wektora cech z warstwy *fc7* oraz w wyniku dostrajania.

		stan fizyczny	
		chory	zdrowy
klasyfikacja	chory	11	10
	zdrowy	13	14

a)

		stan fizyczny	
		chory	zdrowy
klasyfikacja	chory	16	11
	zdrowy	8	13

b)

		stan fizyczny	
		chory	zdrowy
klasyfikacja	chory	15	10
	zdrowy	9	14

c)

Rys. 7.5. Macierze błędów dla klasyfikacji: a) ImageNet – *fc7*, b) ImageNet + KN_19 – *fc7*, c) dostrajanie

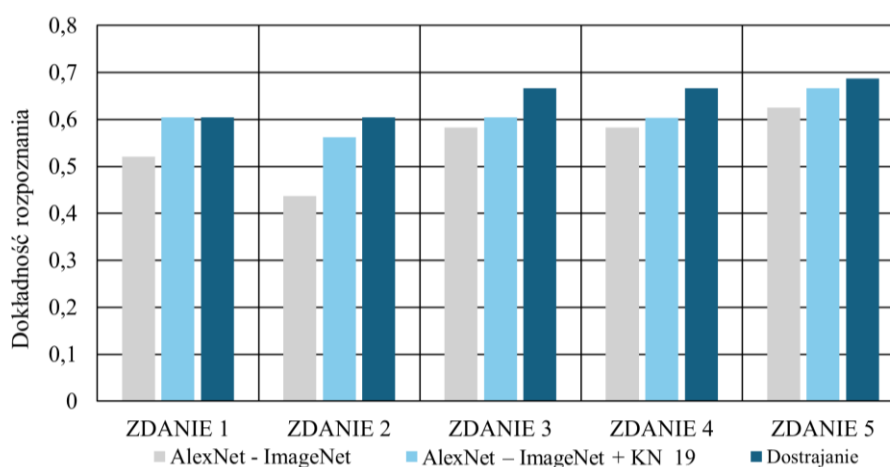
Ze względu na niezadowalające wyniki klasyfikacji w przypadku ekstrakcji wektora cech z warstw *conv5* i *fc6*, kolejne eksperymenty zostały przeprowadzone tylko na podstawie wektora wyodrębnionego z warstwy w pełni połączonej *fc7*.

7.2.2. Rozpoznawanie na podstawie poszczególnych zdań

Kierując się jednym z postulatów niniejszej pracy mówiącym, że zmiany w piśmie wynikające z rozwoju choroby, mogą uwidocznic się dopiero w skutek dłuższego czasu pisania, taki sam eksperyment, jak w podrozdziale 7.2.1, przeprowadzono dla każdego z 5 zdań. Każdorazowo sieć AlexNet została douczona obrazami pisma pochodzącymi z kolejnych zdań. Dzięki temu możliwe było sprawdzenie wpływu czasu pisania na wynik klasyfikacji. Wyniki w postaci dokładności rozpoznania zamieszczono w tabeli 7.4 oraz na rysunku 7.6.

Tabela 7.4. Wyniki dokładności rozpoznania choroby Parkinsona na podstawie poszczególnych zdań

	ekstrakcja cech – AlexNet		dostrajanie
	ImageNet	ImageNet + KN_19	
zdanie 1	0,52 ± 0,21	0,60 ± 0,15	0,60 ± 0,15
zdanie 2	0,44 ± 0,23	0,56 ± 0,20	0,60 ± 0,22
zdanie 3	0,58 ± 0,15	0,60 ± 0,23	0,67 ± 0,18
zdanie 4	0,58 ± 0,20	0,604 ± 0,086	0,67 ± 0,18
zdanie 5	0,62 ± 0,21	0,67 ± 0,25	0,69 ± 0,19



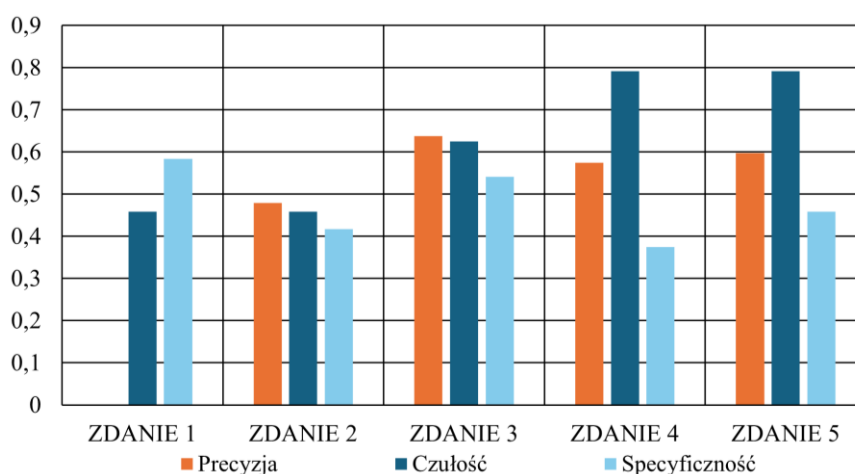
Rys. 7.6. Wykres porównujący dokładność rozpoznania obu klas w zależności od użytego numeru zdania

Analizując powyższe wyniki można zauważyć, że wydłużanie czasu pisania ma wpływ na wynik procesu klasyfikacji. Zarówno dla ekstrakcji cech jak i w przypadku dostrajania, najwyższą dokładność rozpoznania uzyskano dla zdania piątego, a najniższą dla zdania drugiego. Jednocześnie widać tendencję wzrostową między zdaniem drugim a zdaniem piątym. Co ciekawe, również dla wszystkich trzech metod uzyskano identyczny wynik dokładności w przypadku zdania trzeciego i zdania czwartego. Wyższe wyniki dokładności klasyfikacji w przypadku zdania pierwszego w porównaniu do zdania drugiego, mogą wynikać z doświadczenia u pacjentów sytuacji stresowej wynikającej z udziału w badaniu i potrzeby wykonania zapisu w nowych warunkach. Podczas gdy zdanie drugie może odznaczać się większą swobodą wynikającą z zaakceptowania nowej sytuacji, zdenerwowanie podczas zapisu

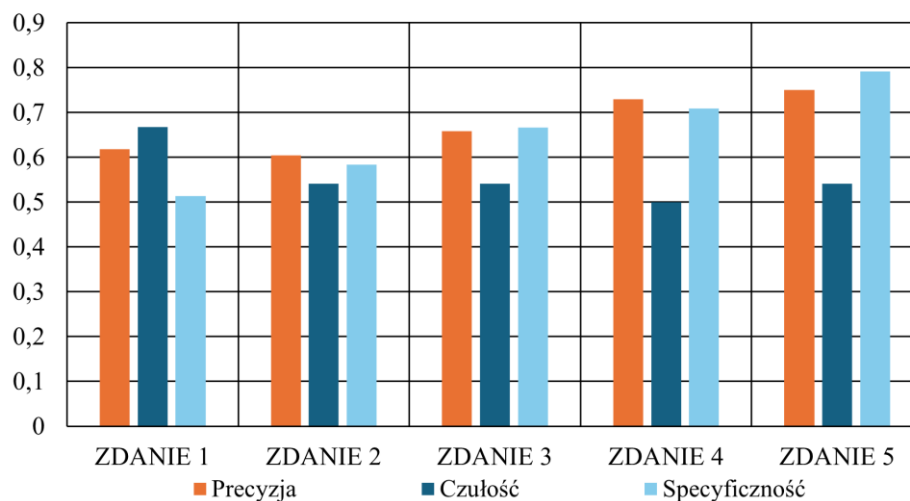
pierwszego zdania mogło spowodować ujawnienie się pewnych cech w piśmie wynikających z choroby i jednocześnie lepszą rozróżnialności grupy chorych i zdrowych. Niezależnie od numeru zdania, najwyższe wyniki dokładności klasyfikacji uzyskano dla podejścia dostrajania, a najniższe dla przypadku ekstrakcji cech z sieci niedouczonej autorską bazą danych KN_19. Najwyższy otrzymany wynik dokładności klasyfikacji równy 69 % uzyskano w przypadku dostrajania z wykorzystaniem obrazów prezentujących zdanie piąte. Pozostałe miary jakości klasyfikacji przedstawiono w tabeli 7.5 i na rysunkach: 7.7, 7.8 i 7.9. Uzyskane macierze błędów przedstawiono na rysunkach 7.10, 7.11 i 7.12.

Tabela 7.5. Miary jakości rozpoznania uzyskane na podstawie poszczególnych zdań

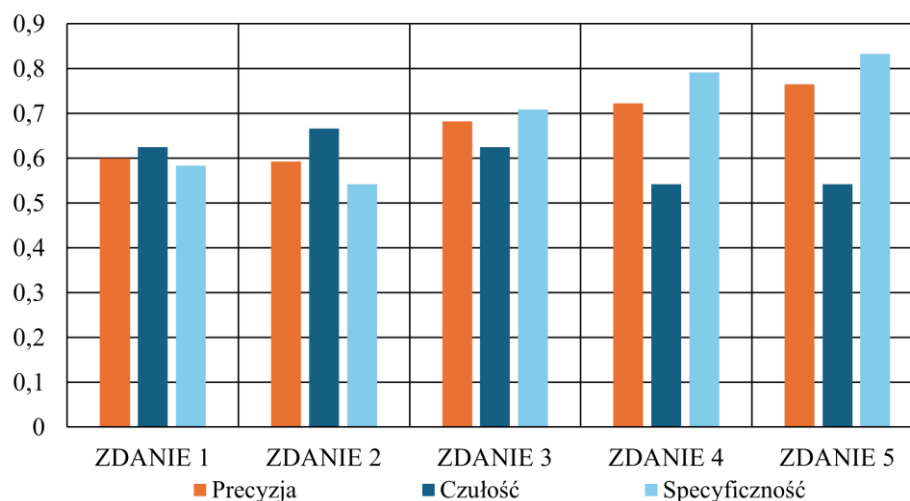
ekstrakcja cech – AlexNet – ImageNet + KN_19					
	zdanie 1	zdanie 2	zdanie 3	zdanie 4	zdanie 5
Dokładność	$0,52 \pm 0,21$	$0,44 \pm 0,23$	$0,58 \pm 0,15$	$0,58 \pm 0,20$	$0,62 \pm 0,21$
Precyzja	-	$0,48 \pm 0,30$	$0,64 \pm 0,24$	$0,58 \pm 0,18$	$0,60 \pm 0,16$
Czułość	$0,46 \pm 0,35$	$0,46 \pm 0,31$	$0,62 \pm 0,21$	$0,79 \pm 0,25$	$0,79 \pm 0,25$
Specyficzność	$0,58 \pm 0,30$	$0,42 \pm 0,34$	$0,54 \pm 0,31$	$0,38 \pm 0,33$	$0,46 \pm 0,25$
ekstrakcja cech – AlexNet – ImageNet + KN_19					
	zdanie 1	zdanie 2	zdanie 3	zdanie 4	zdanie 5
Dokładność	$0,60 \pm 0,15$	$0,56 \pm 0,20$	$0,60 \pm 0,23$	$0,604 \pm 0,086$	$0,67 \pm 0,25$
Precyzja	$0,62 \pm 0,20$	$0,60 \pm 0,22$	$0,66 \pm 0,35$	$0,73 \pm 0,23$	$0,75 \pm 0,38$
Czułość	$0,67 \pm 0,31$	$0,54 \pm 0,25$	$0,54 \pm 0,35$	$0,50 \pm 0,18$	$0,54 \pm 0,25$
Specyficzność	$0,54 \pm 0,25$	$0,58 \pm 0,30$	$0,67 \pm 0,31$	$0,71 \pm 0,28$	$0,79 \pm 0,31$
dostrajanie					
	zdanie 1	zdanie 2	zdanie 3	zdanie 4	zdanie 5
Dokładność	$0,60 \pm 0,15$	$0,60 \pm 0,22$	$0,67 \pm 0,18$	$0,67 \pm 0,18$	$0,69 \pm 0,19$
Precyzja	$0,58 \pm 0,28$	$0,59 \pm 0,17$	$0,74 \pm 0,25$	$0,81 \pm 0,27$	$0,85 \pm 0,27$
Czułość	$0,62 \pm 0,38$	$0,67 \pm 0,25$	$0,62 \pm 0,21$	$0,54 \pm 0,17$	$0,54 \pm 0,17$
Specyficzność	$0,58 \pm 0,24$	$0,54 \pm 0,25$	$0,71 \pm 0,28$	$0,79 \pm 0,31$	$0,83 \pm 0,31$



Rys. 7.7. Precyzja, czułość i specyficzność uzyskane na podstawie ekstrakcji cech z modelu wstępnie przeszkolonego bazą ImageNet dla poszczególnych zdań



Rys. 7.8. Precyzja, czułość i specyficzność uzyskane na podstawie ekstrakcji cech z douczonego, bazą KN_19, modelu dla poszczególnych zdań



Rys. 7.9. Precyzja, czułość i specyficzność uzyskane na podstawie dostrajania modelu dla poszczególnych zdań

		stan fizyczny	
		CH	ZD
klasyfikacja	CH	11	10
	ZD	13	14
a)			
		stan fizyczny	
		CH	ZD
klasyfikacja	CH	11	14
	ZD	13	10
b)			
		stan fizyczny	
		CH	ZD
klasyfikacja	CH	15	11
	ZD	9	13
c)			
		stan fizyczny	
		CH	ZD
klasyfikacja	CH	19	15
	ZD	5	9
d)			
		stan fizyczny	
		CH	ZD
klasyfikacja	CH	19	13
	ZD	5	11
e)			

Rys. 7.10. Macierze błędów otrzymane przy ekstrakcji cech z modelu wstępnie przeszkolonego bazą ImageNet: a) zdanie 1, b) zdanie 2, c) zdanie 3, d) zdanie 4, e) zdanie 5

		stan fizyczny	
		CH	ZD
klasyfikacja	CH	13	8
	ZD	11	16

a)

		stan fizyczny	
		CH	ZD
klasyfikacja	CH	13	10
	ZD	11	14

b)

		stan fizyczny	
		CH	ZD
klasyfikacja	CH	13	8
	ZD	11	16

c)

		stan fizyczny	
		CH	ZD
klasyfikacja	CH	12	7
	ZD	12	17

d)

		stan fizyczny	
		CH	ZD
klasyfikacja	CH	13	5
	ZD	11	19

e)

Rys. 7.11. Macierze błędów otrzymane przy ekstrakcji cech z modelu douczonego bazą KN_19:

a) zdanie 1, b) zdanie 2, c) zdanie 3, d) zdanie 4, e) zdanie 5

		stan fizyczny	
		CH	ZD
klasyfikacja	CH	15	10
	ZD	9	14

a)

		stan fizyczny	
		CH	ZD
klasyfikacja	CH	16	11
	ZD	8	13

b)

		stan fizyczny	
		CH	ZD
klasyfikacja	CH	15	7
	ZD	9	17

c)

		stan fizyczny	
		CH	ZD
klasyfikacja	CH	13	5
	ZD	11	19

d)

		stan fizyczny	
		CH	ZD
klasyfikacja	CH	13	4
	ZD	11	20

e)

Rys. 7.12. Macierze błędów otrzymane przy dostrajaniu modelu: a) zdanie 1, b) zdanie 2, c) zdanie 3, d) zdanie 4, e) zdanie 5

Zauważony wzrost dokładności wraz z każdym kolejnym zdaniem począwszy od zdania drugiego w przypadku ekstrakcji cech z douczonej sieci oraz w przypadku dostrajania jest spowodowany zwiększeniem się liczby poprawnie zdiagnozowanych przypadków zdrowych (TN), co przekłada się na rosnące wartości specyficzności oraz precyzji. Jednocześnie można zauważyć, że wraz z postępem pisanie, liczba poprawnie zdiagnozowanych przypadków chorych (TP), nieznacznie maleje (w przypadku dostrajania) lub jest na stałym poziomie (w przypadku ekstrakcji cech z douczonej sieci). W przypadku ekstrakcji cech z wstępnie przeszkolonej sieci można zauważyć wzrost liczby poprawnie sklasyfikowanych osób chorych (TP) wraz z postępem pisanie, czyli również wzrost czułości. Natomiast nie można wskazać jednoznacznej zależności dla innych miar jakości.

Zaprezentowane w tym rozdziale wyniki jednoznacznie wskazują podejście dostrajania jako najlepsze w kontekście rozwiązywania omawianego problemu. Ze względu na to, kolejne eksperymenty będą przeprowadzane wyłącznie z wykorzystaniem techniki dostrajania.

8. Rozpoznawanie choroby Parkinsona na podstawie sekwencji zdań

W niniejszym rozdziale przeprowadzono klasyfikację binarną osób chorych na chorobę Parkinsona i osób zdrowych na podstawie pisma odręcznego przedstawiającego pojedyncze zdanie, przy czym w tym przypadku w procesie uczenia wykorzystano obrazy przedstawiające wszystkie zdania zapisane przez każdą z osób badanych w ramach tej samej rejestracji. Użycie sekwencji zdań zamiast zdania pojedynczego prowadzi przede wszystkim do zwielokrotnienia liczby danych uczących, co jest zjawiskiem bardzo pożądanym w przypadku systemów uczących. Dodatkowo, jak udowodniono w poprzednim rozdziale, czas trwania procesu pisania pozytywnie wpływa na wynik klasyfikacji. Wynika to ze zmian w piśmie, związanych z obecnością choroby, które mogą pojawić się lub uwypuklić w trakcie wydłużonego wysiłku grafomotorycznego. Wprowadzenie do sieci danych uczących w postaci obrazów prezentujących pełną sekwencje pojedynczych zdań pozwala uwidocznąć te zmiany i wspomóc proces uczenia sieci.

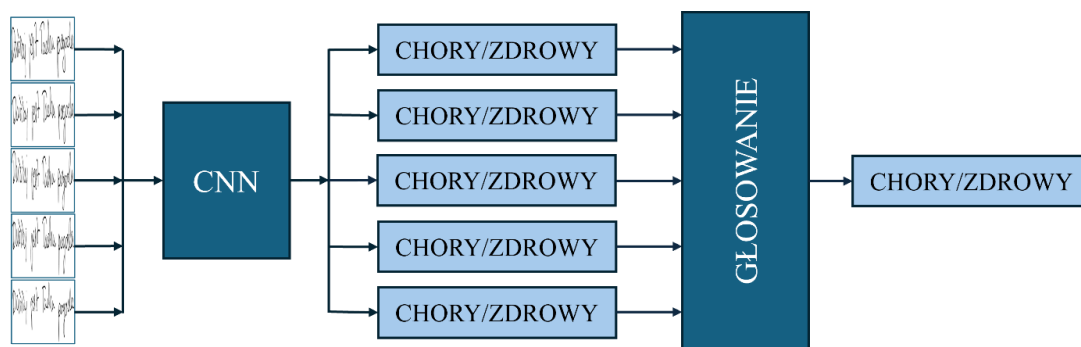
W tym rozdziale przeprowadzono również badanie, w którym zwielokrotnienie danych uczących zrealizowano w procesie klasycznej augmentacji poprzez operacje na obrazie prezentującym pierwsze zdanie, aby ocenić, czy zwiększenie liczby zdań w wyniku przedłużonego pisania przyniosło lepsze rezultaty niż zastosowanie jedynie pojedynczego zdania odpowiednio zwielokrotnionego. Wyniki uzyskane w tym eksperymencie zostały porównane z tymi otrzymanymi na drodze zwielokrotnienia danych poprzez wykorzystanie sekwencji zdań.

8.1. Metodyka badań

8.1.1. Strategia decyzyjna

W tym eksperymencie została wykorzystana technika *transfer learningu* oparta na dostrajaniu z użyciem sieci AlexNet. Dane obrazowe prezentujące wszystkie pojedyncze zdania zapisane podczas rejestracji przez każdą z osób biorących udział w badaniu zostały utworzone w sposób opisany w rozdziale 7.1.2. Następnie, do celów walidacji krzyżowej, zbiór otrzymanych 240 obrazów został podzielony na 8 podzbiorów o jednakowej liczności tak, aby w każdym z nich występowały obrazy prezentujące wszystkie zdania zapisane przez 3 losowo wybrane osoby chore i 3 losowo wybrane osoby zdrowe. Tym sposobem w każdej z 8 iteracji, 210 obrazów zostało użytych do uczenia sieci, a 30 do testowania. Ze względu na fakt, że 5 obrazów testujących pochodzi od jednej osoby, przeprowadzenie procesu klasyfikacji

wymagało opracowania odpowiedniej strategii decyzyjnej. Traktowanie każdego obrazu indywidualnie nie pozwala na pełną ocenę występowania choroby Parkinsona u danej osoby, gdyż 5 zdań dostarcza 5 niezależnych decyzji, które należy zintegrować w jedną spójną decyzję. Wobec powyższego zdecydowano się wygenerować ostateczną decyzję klasyfikacyjną dla danej osoby na podstawie pięciu indywidualnych decyzji otrzymanych na wyjściu klasyfikatora w wyniku kolejnego podania wszystkich pięciu obrazów pochodzących od danej osoby (patrz rys. 8.1).

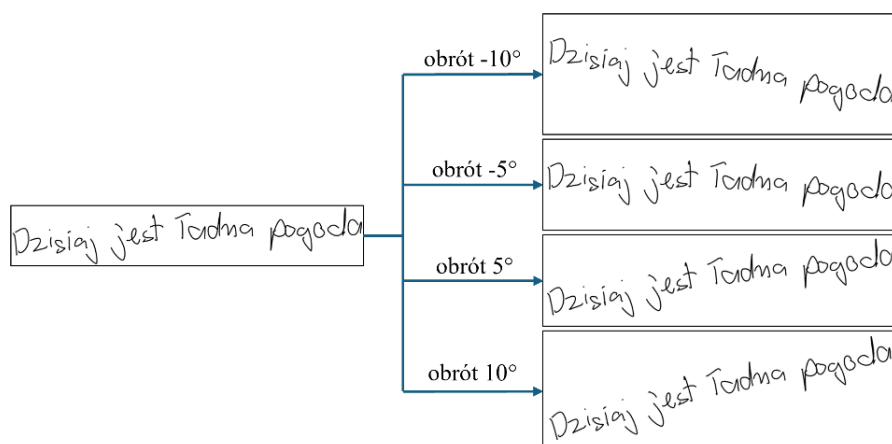


Rys. 8.1. Schemat ogólny podejmowania decyzji klasyfikacyjnej

Naturalnym wyborem strategii wydaje się być głosowanie większościowe (ang. *majority voting*). W tym przypadku oznaczałoby to, że decyzja CHORY zostanie podjęta, gdy minimum 3 zdania pochodzące od danej osoby zostaną sklasyfikowane jako pismo osoby chorej. Zdecydowano się jednak sprawdzić, czy w przypadku rozwiązywanego problemu jest to najskuteczniejsza metoda. Wpływ zmian progu decyzyjnego na ostateczny wynik działania klasyfikatora określano na podstawie krzywej ROC.

8.1.2. Augmentacja pisma odręcznego

W przypadku pisma odręcznego liczba operacji, które można przeprowadzić na obrazie celem wykonania augmentacji jest ograniczona. Obrazy powstałe w wyniku przeprowadzenia takich operacji muszą być odpowiednio reprezentatywne. Zastosowanie odbicia lustrzanego czy obrotu pisma w znacznym zakresie, spowoduje utworzenie danych błędnie ilustrujących pismo odręczne. Również skalowanie czy przesunięcie w przypadku pisma odręcznego prezentującego pełne zdanie spowoduje zmianę wzorca. Jedynym racjonalnym podejściem wydaje się zatem obrót zdania w pewnym niewielkim zakresie. Celem augmentacji w tym przypadku było utworzenie takiej samej liczby danych uczących jak, w przypadku zastosowania sekwencji zdań. Oznaczało to uzyskanie czterech nowych obrazów ze zdania pierwszego. Obrazy wyjściowe były utworzone przez obrót obrazu oryginalnego o kąt: -10° , -5° , 5° , 10° . Nachylenie pisma w takim zakresie jest adekwatne do uzyskiwanego przez człowieka. W ten sposób, podobnie jak poprzednio, uzyskano 240 obrazów, które posłużyły do przeprowadzenia badań. Obrazy powstałe w wyniku augmentacji jednego ze zdań przedstawiono na rysunku 8.2.



Rys. 8.2. Obrazy uczące uzyskane w wyniku obrotu obrazu oryginalnego

8.1.3. Uczenie sieci

Uczenie sieci odbyło się analogicznie jak w rozdziale 7. W związku ze zwiększeniem się zbioru danych, zmieniono tylko rozmiar podzbioru uczącego *mini batch* i ustalono jego wartość na 21, co stanowi równo 1/10 wszystkich przypadków danych uczących w danej walidacji. Zastosowane wartości parametrów uczenia zamieszczono w tabeli 8.1.

Tabela 8.1. Zastosowane wartości podstawowych parametrów uczenia sieci

nazwa parametru (MATLAB)	definicja parametru	wartość
<i>Momentum</i>	współczynnik momentu	0,9
<i>InitialLearnRate</i>	początkowy współczynnik uczenia	0,0001
<i>MiniBatchSize</i>	rozmiar podzbioru uczącego <i>mini batch</i>	21
<i>MaxEpochSize</i>	maksymalna liczba epok uczących	100

8.2. Wyniki eksperymentu

8.2.1. Rozpoznawanie na podstawie zarejestrowanych zdań

W tabeli 8.2. zaprezentowano uzyskane wartości miar jakości klasyfikacji bez zastosowania algorytmu integrowania decyzji. Każde zdanie w tym przypadku traktowane było indywidualnie, jako odrębna próbka. Na rysunku 8.3 zamieszczono macierz błędów dla tego przypadku.

Tabela 8.2. Miary jakości klasyfikacji uzyskane bez głosowania

	ogólne
Dokładność	0,75 ± 0,14
Precyzja	0,77 ± 0,17
Czułość	0,76 ± 0,15
Specyficzność	0,75 ± 0,20

		stan fizyczny	
		chory	zdrowy
klasyfikacja	chory	90	29
	zdrowy	30	91

Rys. 8.3. Macierz błędów uzyskana bez głosowania

Uzyskane wyniki w tym przypadku są zdecydowanie lepsze od wyników klasyfikacji otrzymanych w poprzednim rozdziale. Wszystkie miary jakości klasyfikacji wynoszą około 75 %. Jednak wyniki wypracowane tą metodą wymagają dodatkowego przetworzenia, ponieważ każde z pięciu zdań zapisanych przez tę samą osobę traktowane jest jako odrębna próbka. Aby można było podjąć ostateczną decyzję dotyczącą osoby, konieczne jest zintegrowanie decyzji dotyczących poszczególnych zdań.

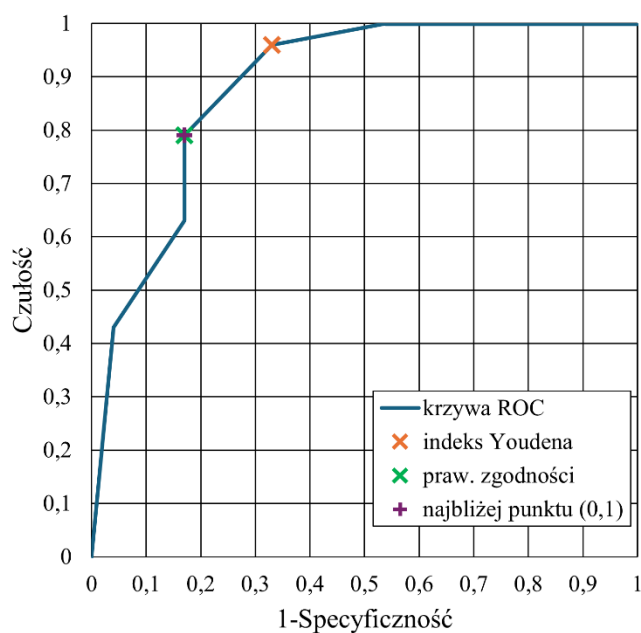
W tabeli 8.3. zaprezentowano wyniki klasyfikacji uzyskane w procesie głosowania przeprowadzonego na podstawie 5 decyzji cząstkowych. Każde zdanie pochodzące od danej osoby podlega niezależnej klasyfikacji, będąc przypisane do odpowiedniej klasy. Następnie zliczana jest liczba przypisań do każdej z klas dla danej osoby. W tabeli 8.3 podany jest próg decyzyjny N zdefiniowany jako minimalna liczba przypisań do grupy chorych, aby ostateczną decyzją sklasyfikować daną osobę jako chorą.

Tabela 8.3. Miary jakości klasyfikacji uzyskane dla różnych wartości progu decyzyjnego

N	decyzja zespołowa				
	1	2	3	4	5
Dokładność	$0,73 \pm 0,15$	$0,81 \pm 0,19$	$0,81 \pm 0,17$	$0,73 \pm 0,18$	$0,69 \pm 0,21$
Precyzja	$0,68 \pm 0,16$	$0,78 \pm 0,20$	$0,86 \pm 0,20$	$0,84 \pm 0,23$	-
Czułość	$1,0 \pm 0,0$	$0,96 \pm 0,12$	$0,79 \pm 0,17$	$0,62 \pm 0,28$	$0,42 \pm 0,34$
Specyficzność	$0,46 \pm 0,31$	$0,67 \pm 0,31$	$0,83 \pm 0,25$	$0,83 \pm 0,25$	$0,96 \pm 0,12$

Zmiana progu decyzyjnego może prowadzić do zwiększenia jednego ze wskaźników wraz z pogorszeniem się innego. Analizując tabelę 8.3 można zauważyć, że zmniejszenie progu decyzyjnego prowadzi do wzrostu czułości, przy jednoczesnym zmniejszeniu wartości specyficzności. Analogicznie wzrost liczby N powoduje zwiększenie parametru specyficzności wraz ze spadkiem czułości. W przypadku N równego 2 i 3 uzyskano największą wartość dokładności klasyfikacji wynoszącą 81 %, przy czym dla N równego 3 wartości miar czułości i specyficzności są bardziej zrównoważone. Spostrzeżenie to potwierdza wykres krzywej ROC zamieszczony na rysunku 8.4 i zaznaczone na nim optymalne punkty odcięcia według trzech kryteriów: punktu będącego najbliższym punktu (0,1), indeksu Youdena oraz prawdopodobieństwa zgodności.

Według kryterium punktu najbliższego punktu (0,1) oraz kryterium prawdopodobieństwa zgodności optymalny punkt odcięcia przypada dla N równego 3, co jest jednoznaczne z zastosowaniem głosowania większościowego. Na N równe 2 jako optymalną wartość progu decyzyjnego wskazało jedynie kryterium Youdena. W związku z powyższym, wyniki uzyskane dla N równego 3 zostały przyjęte jako najkorzystniejsze.



Rys. 8.4. Krzywa ROC klasyfikacji na podstawie obrazów prezentujących sekwencję zdań

Na rysunku 8.5b zamieszczono macierz błędów opracowaną dla tego przypadku zestawioną z najlepszym wynikiem uzyskanym przy wykorzystaniu pojedynczych zdań – rys. 8.5a.

		stan fizyczny	
		chory	zdrowy
klasyfikacja	chory	13	4
	zdrowy	11	20

a)

		stan fizyczny	
		chory	zdrowy
klasyfikacja	chory	19	4
	zdrowy	5	20

b)

Rys. 8.5. Macierze błędów otrzymane przy uczeniu z wykorzystaniem: a) zdania pojedynczego, b) sekwencji zdań

Podsumowując dotychczasowe rozważania można zauważyć, że zastosowanie w procesie uczenia i testowania sieci sekwencji zdań pochodzących z zapisu w ramach jednej rejestracji daje lepsze efekty w postaci wyższych wskaźników jakości klasyfikacji niż użycie zdania pojedynczego. W przypadku użycia tylko pojedynczych zdań od każdej z osób badanych uzyskano maksymalną dokładność rozpoznania choroby Parkinsona wynoszącą 69 %. Zwiększenie liczby danych poprzez dodanie do zbioru obrazów prezentujących wszystkie pięć zdań zapisanych przez każdą z osób oraz zastosowanie odpowiedniej strategii decyzyjnej poprawiło dokładność rozpoznania do wartości 81 %. Poprawa dotyczyła przede wszystkim liczby poprawnie zdiagnozowanych osób chorych (TP), która spowodowała wzrost wartości precyzji (86 %) oraz czułości (79 %). W obu przypadkach uzyskano natomiast identyczną wartość specyficzności, wynoszącą 83 %, chociaż w przypadku sekwencji zdań charakteryzowała się ona mniejszym rozrzutem.

Pozostaje pytanie, czy wzrost dokładności rozpoznania choroby o ponad 12 punktów procentowych, jest spowodowany tylko zwiększeniem liczby danych czy również zmianami w piśmie wynikającymi z zapisu więcej niż jednego zdania. Aby to rozstrzygnąć, przeprowadzono dodatkowy eksperyment, w którym alternatywnie zwielokrotniono liczbę danych uczących poprzez zastosowanie klasycznej augmentacji.

8.2.2. Rozpoznawanie z wykorzystaniem augmentacji danych

W tabeli 8.4 zaprezentowano wyniki rozpoznawania choroby Parkinsona na podstawie danych uzyskanych w wyniku augmentacji. Na rysunku 8.6 zamieszczono macierz błędów dla tego przypadku.

Tabela 8.4. Miary jakości klasyfikacji uzyskane przy zastosowaniu augmentacji

	augmentacja
Dokładność	$0,65 \pm 0,14$
Precyzja	$0,67 \pm 0,21$
Czułość	$0,71 \pm 0,28$
Specyficzność	$0,58 \pm 0,34$

		stan fizyczny	
		chory	zdrowy
klasyfikacja	chory	17	10
	zdrowy	7	14

Rys. 8.6. Macierz błędów uzyskana przy zastosowaniu augmentacji

Klasyczna augmentacja nie zapewniła tak pozytywnych rezultatów, jak w przypadku zastosowania obrazów pochodzących z zapisu sekwencji zdań. Uzyskane wyniki rozpoznania są jednak lepsze od tych otrzymanych w przypadku zdania pojedynczego, gdzie dla zdania 1 dokładność rozpoznania wynosiła 60 %. Na rysunku 8.7 zestawiono macierze błędów otrzymane w wyniku uczenia sieci z wykorzystaniem: zdania pojedynczego, augmentacji zdania pojedynczego i sekwencji zdań.

		stan fizyczny	
		chory	zdrowy
a)	klasyfikacja chory	15	10
	klasyfikacja zdrowy	9	14
b)	klasyfikacja chory	17	10
	klasyfikacja zdrowy	7	14
c)	klasyfikacja chory	19	4
	klasyfikacja zdrowy	5	20

Rys. 8.7. Macierze błędów uzyskane z wykorzystaniem danych uczących w postaci: a) zdania pojedynczego, b) zwielokrotnionego za pomocą augmentacji zdania pojedynczego, c) sekwencji zdań

Analizując zestawienie z rysunku 8.7 można zauważyć, że zastosowanie klasycznej augmentacji w stosunku do jej braku poprawiło tylko rozpoznawalność osób chorych (TP), podczas gdy rozpoznawalność osób zdrowych (TN) pozostała bez zmian. Wykorzystanie sekwencji zdań spowodowało wzrost rozpoznawalności obu grup przez co zdecydowanie

poprawiło wyniki rozpoznania choroby Parkinsona w stosunku do wykorzystania zdania pojedynczego. To spostrzeżenie, oparte na wynikach przeprowadzonych badań w rozdziale siódmym i ósmym, potwierdza pierwszą z postawionych w niniejszej dysertacji tez mówiącą, że użycie sekwencji zdań daje lepsze wyniki klasyfikacji binarnej w diagnostyce choroby Parkinsona niż użycie zdania pojedynczego.

W kolejnym rozdziale przedstawiono klasyfikację binarną osób chorych na chorobę Parkinsona i osób zdrowych na podstawie pisma odręcznego przedstawiającego pojedyncze słowa pochodzące z podziału zdań.

9. Rozpoznawanie choroby Parkinsona na podstawie słów z sekwencji zdań

Spostrzeżenia odnotowane w poprzednich rozdziałach sugerują, że dalsza segmentacja zdań może przynieść pozytywne skutki dla procesu rozpoznawania choroby Parkinsona. Wydaje się, że podział sekwencji zdań na pojedyncze słowa, oprócz naturalnego zwiększenia liczby obrazów zawierających próbki uczące może lepiej niż podział całej rejestracji na zdania odzwierciedlać wpływ choroby na zmiany w piśmie występujące w miarę postępu pisanie. Wszystkie powstałe słowa zostały użyte jako dane uczące zasilające tę samą sieć. Takie podejście nie jest spotykane w literaturze dotyczącej rozwiązywania tego typu problemów, gdzie raczej każdy wzorzec jest podany na odrębną sieć i dopiero ostateczna decyzja jest podejmowana na podstawie wyników uzyskanych przez poszczególne sieci. W niniejszej pracy zdecydowano się na użycie wszystkich słów jednocześnie. Po pierwsze dlatego, aby wspomniany wpływ choroby na pismo był bardziej widoczny, i po drugie, aby system klasyfikacyjny stał się bardziej uniwersalny. W takim przypadku uniezależnia się proces uczenia od treści wzorca, powodując przetwarzanie obrazów przez sieć pod kątem innych zmian niż wynikające z charakteru pisma poszczególnych osób. Zespół graficznych cech pisma jest bardzo indywidualny i zależy od wielu czynników. W przypadku przetwarzania przez sieć tylko jednego słowa, odmienny sposób zapisu znaków pisarskich, może prowadzić do błędnego zdiagnozowania choroby.

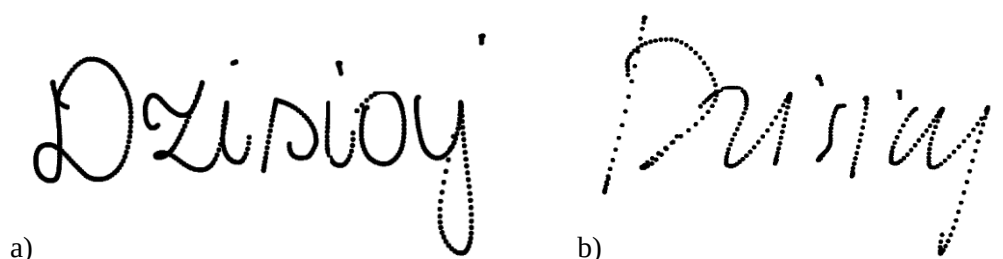
W niniejszym rozdziale w pierwszej kolejności wykazano poprawność tezy mówiącej, że podział sekwencji zdań na sekwencję słów umożliwia dalszą poprawę dokładności rozpoznania choroby Parkinsona na podstawie próbek pisma. Badania w tym kierunku wykonano z uwzględnieniem przesłanek na temat prezentacji danych na obrazie zasugerowanych przez autorów pracy [80]. Po drugie sprawdzono czy sposób prezentacji pisma odręcznego na obrazach użytych do szkolenia sieci konwolucyjnej ma znaczenie, a otrzymane wyniki zestawiono w kontekście przedstawionych w rozprawie tez.

9.1. Metodyka badań

9.1.1. Przygotowanie danych

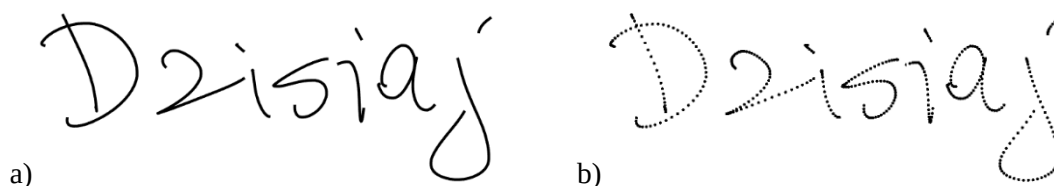
Metoda prezentacji danych na obrazie może mieć istotne znaczenie dla procesu klasyfikacji. W przypadku analizy pisma odręcznego, zastosowanie konwolucyjnych sieci neuronowych, które bazują na gotowych obrazach, wydaje się być mniej efektywne niż zastosowanie

klasycznej parametryzacji pisma. Na statycznych obrazach widoczny jest jedynie ślad zostawiony przez osobę piszącą, bez uwzględnienia aspektów dynamicznych procesu pisania, które można uwypuklić stosując różne techniki inżynierii cech i wykorzystując zaawansowane możliwości pomiarowe tabletu. Jednak odpowiednia rekonstrukcja danych pozwala uwidocznić na obrazie dodatkowe informacje o procesie pisania niż tylko trajektorię ruchu. Do podobnych wniosków doszli autorzy pracy [80], którzy w przeciwieństwie do autorów innych publikacji, na obrazach wykreślili tylko punkty zarejestrowane przez tablet, bez ich interpolacji. Takie podejście wnosi do obrazu informację o prędkości pisania. Ze względu na stałą częstotliwość próbkowania tabletu, gdy prędkość zapisu jest mała, punkty znajdują się bliżej siebie, natomiast gdy prędkość zapisu jest większa, punkty są bardziej odległe. Zilustrowano to na rysunku 9.1, gdzie zaprezentowano dwa obrazy słowa „Dzisiaj” zapisane przez dwie różne osoby w różnym czasie.



Rys. 9.1. Słowo „Dzisiaj” zapisane w czasie: a) 4,7 s, b) 2,4 s

Aby sprawdzić wpływ prezentacji danych na obrazie na wynik klasyfikacji, rozpoznawanie choroby Parkinsona przeprowadzono na obrazach prezentujących interpolowaną trajektorię ruchu rysika oraz z wykorzystaniem samych zarejestrowanych punktów. W przypadku połączonych punktów zastosowano linię o grubości 1 pkt, tak jak miało to miejsce w eksperymentach z pełnymi zdaniami. W przypadku samych punktów zdecydowano się na grubość punktów równą 5 pkt. Tak przyjętą reprezentację graficzną przedstawiono na rysunku 9.2.



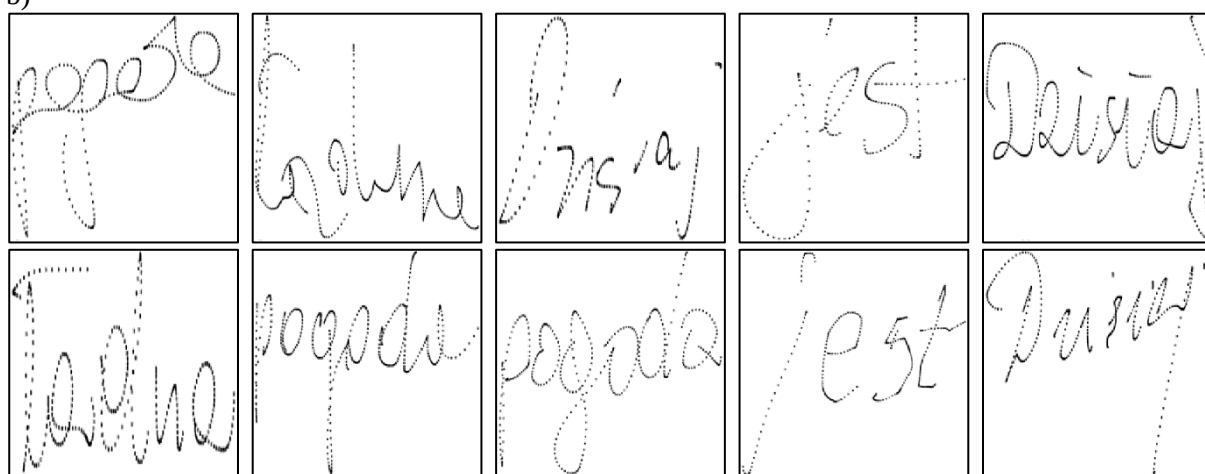
Rys. 9.2. Reprezentacja danych na obrazie z zastosowaniem: a) interpolacji, b) punktów

Wykreślone obrazy prezentujące pojedyncze słowa, następnie przewymiarowano tak, aby ich rozmiar odpowiadał wymaganiom przez sieć AlexNet – $227 \times 227 \times 3$. Przykładowe obrazy zaprezentowano na rysunku 9.3.

a)



b)



Rys. 9.3. Przykładowe obrazy pojedynczych słów użyte jako dane wejściowe sieci, wykreślone z zastosowaniem: a) interpolacji, b) samych punktów

W wyniku podziału zdań na słowa otrzymano 960 obrazów. Uzyskany zbiór obrazów podzielono na 8 równych podzbiorów, według osób, w taki sposób aby w jednej iteracji walidacyjnej wśród danych testujących występowały obrazy prezentujące wszystkie słowa zapisane przez 3 losowo wybrane osoby chore i 3 losowo wybrane osoby zdrowe. W każdej z 8 iteracji, 840 obrazów zostało użytych do uczenia sieci, a 120 do testowania. Analogicznie jak w przypadku zdań, ostateczna decyzja klasyfikacyjna dla jednej osoby była podejmowana na podstawie dwudziestu indywidualnych decyzji otrzymanych na wyjściu klasyfikatora w wyniku przetwarzania wszystkich dwudziestu obrazów pochodzących od danej osoby. Optymalny próg decyzyjny określono na podstawie krzywej ROC.

9.1.2. Uczenie sieci

Uczenie sieci odbyło się analogicznie jak w poprzednich eksperymentach. Również rozmiar podzbioru uczącego *mini batch* pozostał bez zmian i wynosił 21, co stanowiło w tym przypadku 1/40 wszystkich danych uczących w danej iteracji. Zastosowane wartości parametrów uczenia zamieszczono w tabeli 9.1.

Tabela 9.1. Zastosowane wartości podstawowych parametrów uczenia sieci

nazwa parametru (MATLAB)	definicja parametru	wartość
<i>Momentum</i>	współczynnik momentu	0,9
<i>InitialLearnRate</i>	początkowy współczynnik uczenia	0,0001
<i>MiniBatchSize</i>	rozmiar podzbioru uczącego <i>mini batch</i>	21
<i>MaxEpochSize</i>	maksymalna liczba epok uczących	100

9.2. Wyniki eksperymentu

9.2.1. Rozpoznawanie na podstawie sekwencji słów

Poniżej zaprezentowano wartości miar jakości klasyfikacji w przypadku danych obrazowych prezentujących pojedyncze słowa w postaci: połączonych punktów (tabela 9.2) oraz samych punktów (tabela 9.3) dla kolejnych wartości progu decyzyjnego N , który zdefiniowano jako minimalną liczbę przypisań do grupy chorych, aby ostateczną decyzją sklasyfikować daną osobę jako chorą.

Tabela 9.2. Miary jakości klasyfikacji uzyskane dla różnych wartości progu decyzyjnego dla obrazów w postaci połączonych punktów

decyzja zespołowa										
N	1	2	3	4	5	6	7	8	9	10
<i>Dokładność</i>	0,54	0,54	0,54	0,52	0,60	0,67	0,73	0,75	0,79	0,85
<i>Precyzja</i>	0,53	0,53	0,53	0,51	0,57	0,62	0,69	0,73	0,78	0,86
<i>Czułość</i>	1,0	0,96	0,96	0,92	0,92	0,92	0,92	0,88	0,88	0,88
<i>Specyficzność</i>	0,083	0,12	0,12	0,12	0,29	0,42	0,54	0,62	0,71	0,83

decyzja zespołowa										
N	11	12	13	14	15	16	17	18	19	20
<i>Dokładność</i>	0,77	0,73	0,69	0,65	0,65	0,62	0,58	0,54	0,52	0,50
<i>Precyzja</i>	0,84	0,84	0,84	-	-	-	-	-	-	-
<i>Czułość</i>	0,71	0,54	0,46	0,38	0,33	0,25	0,17	0,083	0,042	0,00
<i>Specyficzność</i>	0,83	0,92	0,92	0,92	0,96	1,0	1,0	1,0	1,0	1,0

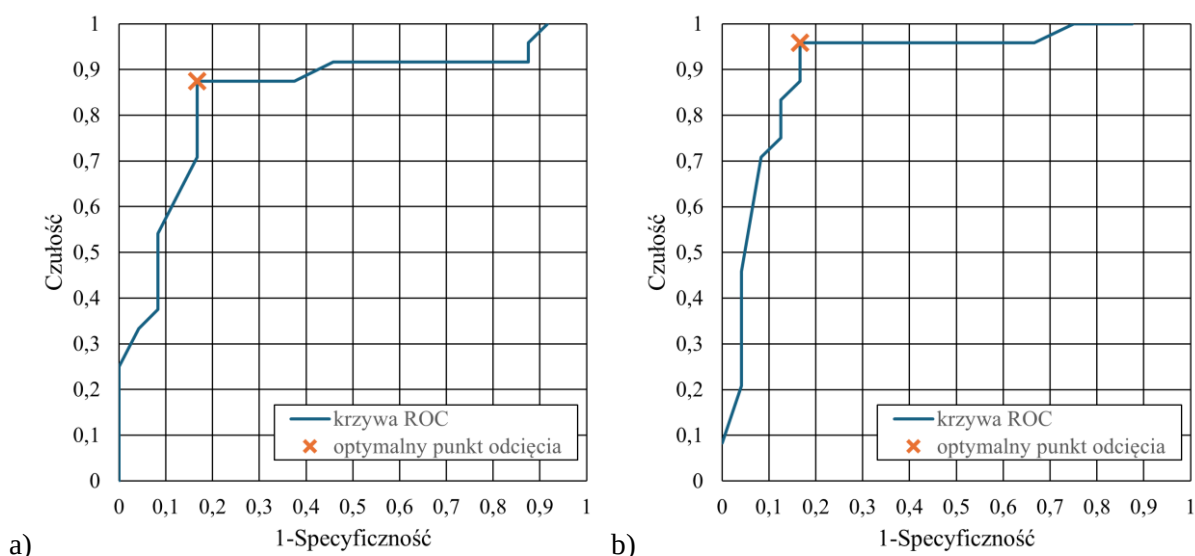
Tabela 9.3. Miary jakości klasyfikacji uzyskane dla różnych wartości progu decyzyjnego dla obrazów w postaci punktów

decyzja zespołowa										
N	1	2	3	4	5	6	7	8	9	10
<i>Dokładność</i>	0,56	0,56	0,62	0,65	0,73	0,75	0,79	0,81	0,83	0,90
<i>Precyzja</i>	0,54	0,54	0,58	0,60	0,71	0,72	0,78	0,79	0,81	0,88
<i>Czułość</i>	1,00	1,00	1,00	0,96	0,96	0,96	0,96	0,96	0,96	0,96
<i>Specyficzność</i>	0,12	0,12	0,25	0,33	0,50	0,54	0,62	0,67	0,71	0,83

Tabela 9.3. Miary jakości klasyfikacji uzyskane dla różnych wartości progu decyzyjnego dla obrazów w postaci punktów (cd.)

decyzja zespołowa										
<i>N</i>	11	12	13	14	15	16	17	18	19	20
<i>Dokładność</i>	0,85	0,85	0,83	0,83	0,81	0,81	0,71	0,62	0,58	0,54
<i>Precyzja</i>	0,85	0,90	0,90	0,90	0,90	0,93	-	-	-	-
<i>Czułość</i>	0,88	0,83	0,79	0,79	0,75	0,71	0,46	0,29	0,21	0,083
<i>Specyficzność</i>	0,83	0,88	0,88	0,88	0,88	0,92	0,96	0,96	0,96	1,0

Podobnie jak w przypadku sekwencji zdań, zwiększenie progu decyzyjnego powoduje wzrost specyficzności przy jednoczesnym spadku czułości. Na podstawie wyników zawartych w tabelach 9.2 i 9.3 widać, że dla obu typów danych najlepszy rezultat rozpoznawania choroby uzyskano dla N równego 10, co nie jest równoznaczne z zastosowaniem głosowania większościowego. W tradycyjnej odmianie trybu głosowania większościowego jest przyjęcie założenia, że zwycięzca musi uzyskać minimum 50 % plus jeden głos. W przeciwnym przypadku przyjmuje się brak rozstrzygnięcia [95].



Rys. 9.4. Krzywa ROC klasyfikacji na podstawie obrazów słów wykreślonych z użyciem:
a) interpolacji, b) punktów

Na rysunku 9.4 zamieszczono krzywe ROC dla obu analizowanych przypadków, z zaznaczonym optymalnym punktem odcięcia wyznaczonym z wykorzystaniem trzech metod optymalizacyjnych: indeksu Youdena, punktu najbliższego punktu (0,1) oraz prawdopodobieństwa zgodności. Wszystkie trzy kryteria optymalizacyjne wskazały punkt odpowiadający N równemu 10, jako optymalny próg decyzyjny. W tabeli 9.4 zamieszczono miary jakości klasyfikacji otrzymane dla dwóch rozważanych metod prezentacji danych na obrazie wraz z wartościami odchyleń standardowych dla optymalnego progu decyzyjnego.

Tabela 9.4. Miary jakości klasyfikacji uzyskane dla optymalnego progu decyzyjnego dla obu przypadków prezentacji danych na obrazie.

	interpolacja	punkty
Dokładność	$0,85 \pm 0,11$	$0,896 \pm 0,086$
Precyzja	$0,86 \pm 0,15$	$0,88 \pm 0,13$
Czułość	$0,88 \pm 0,17$	$0,96 \pm 0,12$
Specyficzność	$0,83 \pm 0,18$	$0,83 \pm 0,18$

Jak widać metoda prezentacji danych na obrazie w przypadku klasyfikacji z użyciem słów ma znaczenie. Uzyskano wyższe wartości wszystkich miar jakości w przypadku, gdy na obrazach użytych do przetwarzania przez sieć wykreślono tylko zarejestrowane przez tablet punkty. Wprowadzenie interpolacji tych punktów w taki sposób, aby na obrazie widoczny był pełny zapisany przez pacjenta ślad nie wpływa pozytywnie na wynik klasyfikacji, gdzie dokładność rozpoznania wyniosła 85 %. Natomiast zrekonstruowanie obrazu pisma tylko za pomocą punktów spowodowało uzyskanie dokładności rozpoznania choroby równej 89,6 %. W obu przypadkach czułość przewyższa wartość specyficzności, co oznacza, że klasyfikatory charakteryzują się lepszą rozpoznawalnością osób chorych (TP) niż osób zdrowych (TN). Można to zaobserwować analizując macierze błędów zaprezentowane na rysunku 9.5.

		stan fizyczny	
		chory	zdrowy
klasyfikacja	chory	21	4
	zdrowy	3	20

a)

		stan fizyczny	
		chory	zdrowy
klasyfikacja	chory	23	4
	zdrowy	1	20

b)

Rys. 9.5. Macierze błędów uzyskane na podstawie obrazów słów w postaci: a) połączonych punktów, b) tylko zarejestrowanych punktów

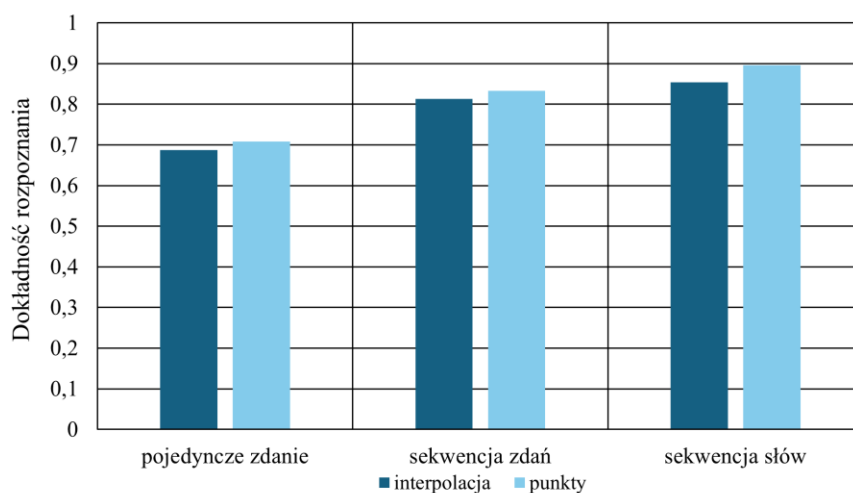
Analizując wyniki tego badania można zauważyć, że niezależnie od prezentacji danych na obrazie, uzyskano wyniki rozpoznania choroby Parkinsona wyższe niż w przypadku zastosowania sekwencji zdań, gdzie najwyższa otrzymana dokładność wynosiła 81 %. Jest to wniosek, który potwierdza drugą postawioną w tej rozprawie tezę, że możliwa jest poprawa dokładności klasyfikacji binarnej w diagnostyce choroby Parkinsona na drodze podziału sekwencji zdań na sekwencję słów. Pewne wątpliwości może budzić fakt, że w przypadku analizy słów skupiono się na obrazach bez interpolacji a w przypadkach poprzednich, które były swoistą polemiką z danymi literaturowymi, analizowano używany w nich typ obrazów z interpolacją. Innymi słowy powstaje pytanie, czy sam fakt użycia danych bez interpolacji nie jest czynnikiem, który w największym stopniu wpływa na uzyskiwane wyniki. Aby to sprawdzić przeprowadzono również serię badań z wykorzystaniem pojedynczego zdania i sekwencji zdań z wykorzystaniem samych punktów.

9.2.2. Wpływ prezentacji danych na obrazie na wynik klasyfikacji – porównanie wyników

W tabeli 9.5 oraz na rysunku 9.6 zestawiono ze sobą wyniki dokładności rozpoznania choroby Parkinsona uzyskane na podstawie obrazów prezentujących pismo odręczne w postaci: pojedynczych zdań, pojedynczych zdań pochodzących z zapisu ich sekwencji oraz pojedynczych słów pochodzących z podziału zdań na słowa dla dwóch metod prezentacji danych na obrazie – z interpolacją i bez.

Tabela 9.5. Wyniki dokładności rozpoznania choroby Parkinsona dla dwóch metod prezentacji danych na obrazie i trzech typów danych

	pojedyncze zdanie	sekwencja zdań	sekwencja słów
interpolacja	$0,69 \pm 0,19$	$0,81 \pm 0,17$	$0,85 \pm 0,11$
punkty	$0,71 \pm 0,15$	$0,83 \pm 0,13$	$0,896 \pm 0,086$



Rys. 9.6. Wykres porównujący dokładność rozpoznania obu klas w zależności od użytej metody prezentacji danych na obrazie i typu danych

Na podstawie powyższego porównania widać, że niezależnie od sposobu prezentacji danych na obrazie zmiana typu danych z obrazów prezentujących zdanie pojedyncze na obrazy prezentujące zdania pochodzące z zapisu ich sekwencji i dalej na obrazy prezentujące słowa powstałe na drodze podziału sekwencji zdań prowadzi do poprawy wyników dokładności rozpoznania choroby Parkinsona. Jednocześnie zmiana typu danych prowadzi do uzyskania korzystniejszych wyników niż zmiana prezentacji danych na obrazie. Użycie sekwencji zdań zamiast zdania pojedynczego poprawia dokładność klasyfikacji o 12 punktów procentowych. Użycie sekwencji słów zamiast sekwencji zdań w mniejszym stopniu poprawia dokładność rozpoznania niż zamiana zdania pojedynczego na sekwencję zdań. Dla danych w postaci połączonych punktów nastąpił wzrost dokładności o 4 punkty procentowe, dla danych w postaci samych punktów o 7 punktów procentowych. Najlepszy wynik rozpoznania choroby Parkinsona uzyskano dla danych w postaci słów pochodzących z zapisu sekwencji zdań,

zaprezentowanych na obrazie w postaci samych punktów. Dokładność klasyfikacji w tym przypadku wynosiła 89,6 %. W tabeli 9.6 zamieszczono zestawienie pozostałych miar jakości klasyfikacji. Na rysunkach 9.7 i 9.8 zamieszczono zestawienie macierzy błędów dla wszystkich rozpatrywanych przypadków.

Tabela 9.6. Miary jakości klasyfikacji dla dwóch metod prezentacji danych na obrazie i trzech typów danych

interpolacja			
	<i>pojedyncze zdanie</i>	<i>sekwencja zdań</i>	<i>sekwencja słów</i>
Dokładność	0,69 ± 0,19	0,81 ± 0,17	0,85 ± 0,11
Precyzja	0,85 ± 0,27	0,86 ± 0,20	0,85 ± 0,15
Czułość	0,54 ± 0,17	0,79 ± 0,17	0,88 ± 0,17
Specyficzność	0,83 ± 0,31	0,83 ± 0,25	0,83 ± 0,18
punkty			
	<i>pojedyncze zdanie</i>	<i>sekwencja zdań</i>	<i>sekwencja słów</i>
Dokładność	0,71 ± 0,15	0,83 ± 0,13	0,896 ± 0,086
Precyzja	0,75 ± 0,22	0,85 ± 0,16	0,88 ± 0,13
Czułość	0,71 ± 0,21	0,83 ± 0,18	0,96 ± 0,12
Specyficzność	0,71 ± 0,28	0,83 ± 0,18	0,83 ± 0,18

		stan fizyczny	
		chory	zdrowy
klasyfikacja	chory	13	4
	zdrowy	11	20

a)

		stan fizyczny	
		chory	zdrowy
klasyfikacja	chory	19	4
	zdrowy	5	20

b)

		stan fizyczny	
		chory	zdrowy
klasyfikacja	chory	21	4
	zdrowy	3	20

c)

Rys. 9.7. Macierze błędów uzyskane dla obrazów w postaci połączonych punktów z wykorzystaniem: a) zdania pojedynczego, b) sekwencji zdań, c) sekwencji słów

		stan fizyczny	
		chory	zdrowy
klasyfikacja	chory	17	7
	zdrowy	7	17

a)

		stan fizyczny	
		chory	zdrowy
klasyfikacja	chory	20	4
	zdrowy	4	20

b)

		stan fizyczny	
		chory	zdrowy
klasyfikacja	chory	23	4
	zdrowy	1	20

c)

Rys. 9.8. Macierze błędów uzyskane dla obrazów w postaci punktów z wykorzystaniem: a) zdania pojedynczego, b) sekwencji zdań, c) sekwencji słów

W przypadku połączonych punktów na obrazie, wzrost dokładności klasyfikacji jest wywołany zwiększeniem się liczby poprawnie sklasyfikowanych przypadków chorych (TP),

podczas gdy liczba prawidłowo sklasyfikowanych przypadków zdrowych (TN) pozostaje bez zmian i wynosi 20. Podobnie w przypadku zmiany metody prezentacji danych na obrazie, między użyciem sekwencji zdań a sekwencji słów nastąpił wzrost liczby prawidłowo sklasyfikowanych przypadków chorych (TP), podczas gdy liczba prawidłowo sklasyfikowanych przypadków zdrowych (TN) również pozostaje bez zmian i wynosi 20. Natomiast w przypadku użycia sekwencji zdań zamiast zdania pojedynczego poprawie uległo, w równym stopniu, rozpoznanie osób chorych jak i osób zdrowych.

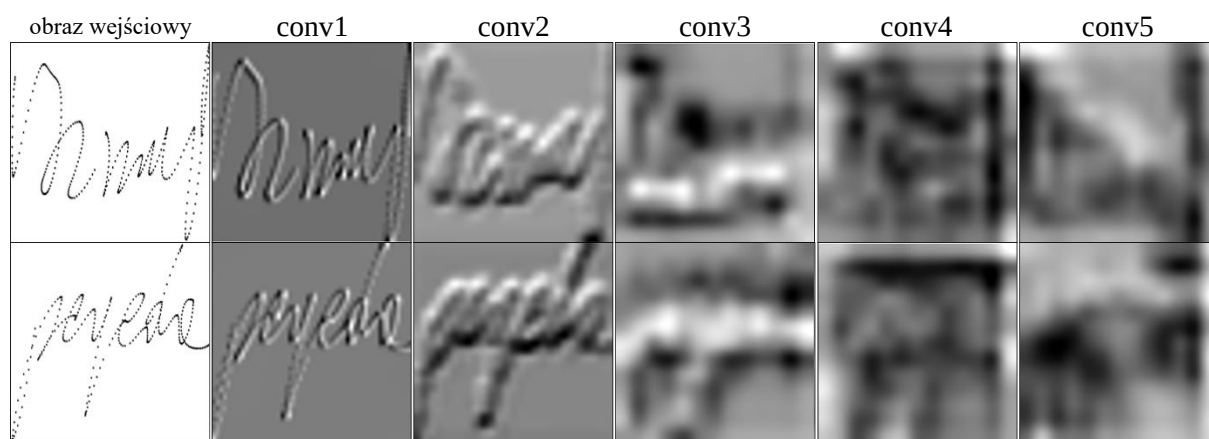
10. Prace optymalizacyjne w zakresie doboru klasyfikatora

W badaniach przeprowadzonych w rozdziałach: ósmym i dziewiątym, skupiono się na poprawie wyników klasyfikacji poprzez zwiększenie liczby danych zasilających standardową sieć konwolucyjną, za jaką można uważać AlexNet na drodze podziału pełnej rejestracji pisma zawierającej pięć jednakowych zdań, najpierw na pojedyncze zdania, a następnie na słowa. Wykazano, że zastosowanie danych uczących w postaci obrazów prezentujących pojedyncze słowa daje najlepsze rezultaty i przy odpowiedniej metodzie prezentacji danych na obrazie oraz wykorzystaniu sieci AlexNet w podstawowym trybie dostrajania wszystkich warstw, umożliwia uzyskanie dokładności rozpoznania równej 89,6 %. Zaprezentowany materiał nie wyklucza jednak możliwości dalszej poprawy tego wyniku na drodze prac mających na celu dobór struktury sieci odpowiedniej do charakteru wykorzystywanych danych. Niniejszy rozdział przedstawia dwa kierunki badań przeprowadzonych w celu realizacji tego zadania – modyfikację procesu dostrajania sieci AlexNet i zastosowanie innych, bardziej zaawansowanych struktur sieci.

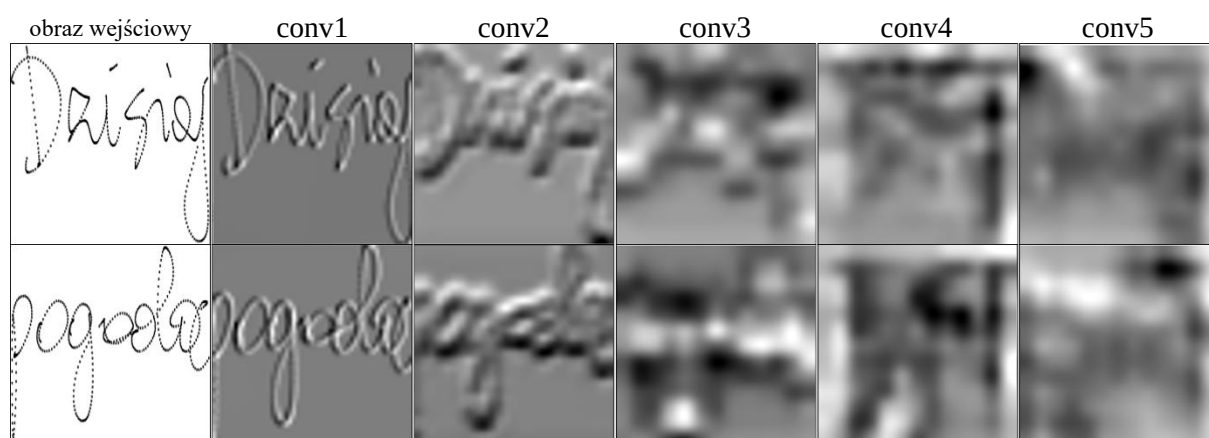
10.1. Dobór głębokości dostrojenia sieci AlexNet

We wszystkich zaprezentowanych dotychczas w niniejszej rozprawie eksperymentach, zastosowano identyczną technikę douczania sieci, w której wszystkie warstwy sieci podlegały procesowi aktualizacji wag. Na etapie projektowania badań eksperymentalnych i wykonywania pierwszych prób uczenia sieci za pomocą danych prezentujących pismo odręczne, takie podejście wydawało się być najkorzystniejsze. Wiązało się to ze specyfiką rozwiązywanego problemu oraz wyraźnie innym charakterem danych prezentujących pismo odręczne niż obrazów z bazy ImageNet, którymi sieć została wstępnie wytrenowana. Dostępne w literaturze wyniki badań wskazują, że douczanie tylko pewnej liczby warstw konwolucyjnych, przy zamrożeniu pozostałych może poprawić wydajność sieci prowadząc do poprawy wyników klasyfikacji [90, 103, 104]. Jednak wskazanie granicy między warstwami mającymi przejść proces douczenia a warstwami zamrożonymi, pozwalającej uzyskać najlepszy wynik, jest zależne od stopnia podobieństwa danych zadania źródłowego i zadania docelowego oraz od poziomu szczegółowości, pozwalającego zaklasyfikować dany przypadek do odpowiedniej grupy [103]. Wynika to z faktu, że sieci konwolucyjne konstruuja hierarchiczną reprezentację obrazów wejściowych. Początkowe warstwy mają za zadanie wyławiać pewne charakterystyczne elementy obrazu, tak zwane cechy niskiego poziomu, jak grupy pikseli tworzące określony kształt lub krawędzie ułożone w różnych pozycjach. W głębszych

warstwach generowane są bardziej szczegółowe cechy obrazu, adekwatne do rozwiązywanego zadania. W przypadku pisma odręcznego, początkowe warstwy będą eksponować różnice wynikające na przykład z odmiennego charakteru pisma czy innego wzorca, w przypadku gdy sieć jest zasilana obrazami różnych słów. Natomiast różnice wynikające z obecności choroby, mogą być na tyle dyskretne i niejednoznaczne, że staną się zauważalne dopiero w wyniku przetwarzania przez głębsze warstwy sieci. Rysunek 10.1 przedstawia mapy cech z losowo wybranych kanałów aktywacji na wyjściach poszczególnych warstw konwolucyjnych dla dwóch przypadków obrazów wejściowych przedstawiających różne słowa zapisane przez osobę chorą. Analogicznie na rysunku 10.2 przedstawiono przykładowe mapy cech generowane w poszczególnych warstwach konwolucyjnych na podstawie obrazów wejściowych zapisanych przez osobę zdrową. W przypadku każdej z warstw dla poszczególnych słów i osób zaprezentowane obrazy pochodziły z tych samych kanałów aktywacji.



Rys. 10.1. Przykładowe obrazy wygenerowane przez poszczególne warstwy konwolucyjne na podstawie danych pochodzących od osoby chorej



Rys. 10.2. Przykładowe obrazy wygenerowane przez poszczególne warstwy konwolucyjne na podstawie danych pochodzących od osoby zdrowej

Na obrazach zaprezentowanych na rysunkach 10.1 i 10.2, widać, że przetwarzanie przez początkowe warstwy sieci uwypukla różnice między poszczególnymi wzorcami słów. Im

głębsza warstwa tym różnice te stają się mniej widoczne, a sieć generuje cechy o wyższym poziomie szczegółowości. Na wizualizacjach działania głębszych warstw, zauważalne stają się różnice między pismem osoby chorej i osoby zdrowej, bez względu na treść zapisanego słowa.

W badaniach zmierzających do oceny, czy możliwa jest poprawa wyniku rozpoznania choroby poprzez zastosowanie zamrażania pewnej liczby warstw sieci AlexNet i douczania pozostałych rozpoczęto od zamrożenia pierwszej warstwy konwolucyjnej, a następnie kolejno zamrażano następne warstwy, przeprowadzając dla każdego przypadku proces uczenia i testowania sieci z wykorzystaniem dotychczasowej metodyki. W tabeli 10.1 zamieszczono nazwy warstw, które zostały zamrożone i które podlegały procesowi douczania w poszczególnych eksperymentach wraz z identyfikatorami powstałych tym sposobem sieci.

Tabela 10.1. Wykaz zamrożonych i douczanych warstw w poszczególnych eksperymentach

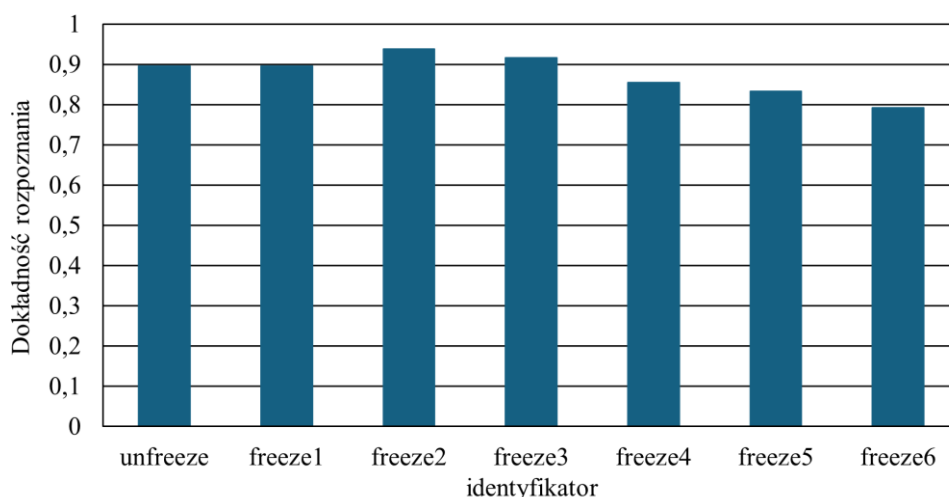
sieć	zamrożone warstwy	warstwy podlegające uczeniu
<i>unfreeze</i>	-	conv1, conv2, conv3, conv4, conv5, fc6, fc7, klasyfikująca
<i>freeze1</i>	conv1	conv2, conv3, conv4, conv5, fc6, fc7, klasyfikująca
<i>freeze2</i>	conv1, conv2	conv3, conv4, conv5, fc6, fc7, klasyfikująca
<i>freeze3</i>	conv1, conv2, conv3	conv4, conv5, fc6, fc7, klasyfikująca
<i>freeze4</i>	conv1, conv2, conv3, conv4	conv5, fc6, fc7, klasyfikująca
<i>freeze5</i>	conv1, conv2, conv3, conv4, conv5	fc6, fc7, klasyfikująca
<i>freeze6</i>	conv1, conv2, conv3, conv4, conv5, fc6	fc7, klasyfikująca

W tabeli 10.2. zaprezentowano miary jakości klasyfikacji uzyskane dla sieci powstałych z tytułu zmiany głębokości dostrajania.

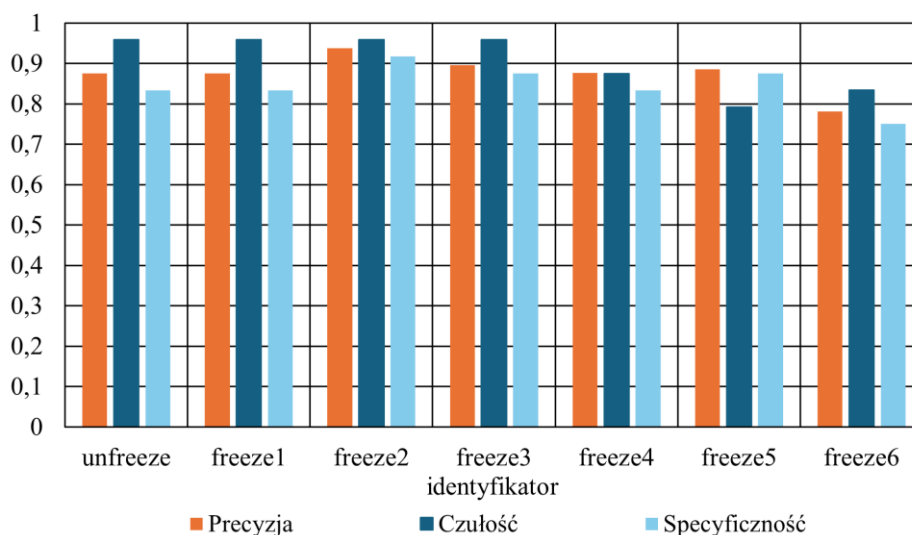
Tabela 10.2. Miary jakości klasyfikacji uzyskane dla sieci o różnych głębokościach dostrojenia wag

sieć	<i>unfreeze</i>	<i>freeze1</i>	<i>freeze2</i>	<i>freeze3</i>
<i>Dokładność</i>	0,896 ± 0,086	0,896 ± 0,086	0,938 ± 0,086	0,92 ± 0,13
<i>Precyzja</i>	0,88 ± 0,13	0,88 ± 0,13	0,94 ± 0,12	0,90 ± 0,15
<i>Czułość</i>	0,96 ± 0,12	0,96 ± 0,12	0,96 ± 0,12	0,96 ± 0,12
<i>Specyficzność</i>	0,83 ± 0,18	0,83 ± 0,18	0,92 ± 0,16	0,88 ± 0,17
sieć	<i>freeze4</i>	<i>freeze5</i>	<i>freeze6</i>	
<i>Dokładność</i>	0,85 ± 0,14	0,83 ± 0,13	0,79 ± 0,12	
<i>Precyzja</i>	0,88 ± 0,17	0,89 ± 0,16	0,78 ± 0,14	
<i>Czułość</i>	0,88 ± 0,17	0,79 ± 0,17	0,83 ± 0,18	
<i>Specyficzność</i>	0,83 ± 0,25	0,88 ± 0,17	0,75 ± 0,15	

Miary jakości klasyfikacji zestawiono w postaci wykresów słupkowych i zamieszczono na rysunkach 10.3 i 10.4.



Rys. 10.3. Dokładność klasyfikacji w zależności od głębokości douczania sieci



Rys. 10.4. Precyzja, czułość i specyficzność uzyskane dla różnych głębokości douczania sieci

Analizując wyniki z tabeli 10.2 oraz wykresy 10.3 i 10.4 można zauważyć, że głębokość douczania sieci konwolucyjnej AlexNet ma wpływ na ostateczny wynik klasyfikacji. Wyniki uzyskane dla poszczególnych głębokości dostrajania sieci są do siebie zbliżone, jednak najlepszy rezultat uzyskano w przypadku zamrożenia pierwszej i drugiej warstwy konwolucyjnej, douczając strukturę sieci bazą danych KN_19 od trzeciej warstwy konwolucyjnej. Tą metodą uzyskano wynik dokładności rozpoznania choroby równy 93,8 %, co jest najlepszym wynikiem uzyskanym przez autora niniejszej rozprawy. Zamrażając coraz większą liczbę warstw uzyskiwano gorszy wynik dokładności rozpoznania. Na rysunku 10.5. przedstawiono uzyskane macierze błędów.

		stan fizyczny	
		chory	zdrowy
klasyfikacja	chory	23	4
	zdrowy	1	20

a)

		stan fizyczny	
		chory	zdrowy
klasyfikacja	chory	23	2
	zdrowy	1	22

b)

		stan fizyczny	
		chory	zdrowy
klasyfikacja	chory	23	3
	zdrowy	1	21

c)

		stan fizyczny	
		chory	zdrowy
klasyfikacja	chory	21	4
	zdrowy	3	20

d)

		stan fizyczny	
		chory	zdrowy
klasyfikacja	chory	19	3
	zdrowy	5	21

e)

		stan fizyczny	
		chory	zdrowy
klasyfikacja	chory	20	6
	zdrowy	4	18

f)

Rys. 10.5. Macierze błędów uzyskane dla przypadku: a) freeze1, b) freeze2, c) freeze3, d) freeze4, e) freeze5, f) freeze6

Zamrożenie początkowych warstw sieci (przypadek: freeze1, freeze2 i freeze3) powodowało bardzo dobrą rozróżnialność osób chorych (TP), co przekładało się na uzyskany wynik czułości równy 96 %. Zamrażanie coraz większej liczby warstw spowodowało spadek wartości wszystkich pozostałych miar jakości klasyfikacji: czułości, precyzji i specyficzności.

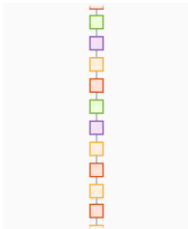
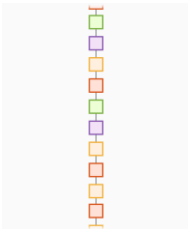
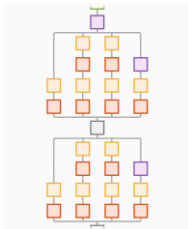
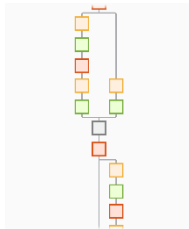
10.2. Wykorzystanie innych sieci konwoucyjnych

Rozwój sieci konwoucyjnych, po sukcesie AlexNet na olimpiadzie ILSVRC w 2012 r. [102], doprowadził do sytuacji, w której do rozwiązywania zadań klasyfikacji na podstawie danych obrazowych możliwe są do wykorzystania również inne, bardziej zaawansowane struktury. W niniejszej pracy, na bazie proponowanej strategii przygotowania danych obrazowych, zrealizowano badania porównawcze z wykorzystaniem sieci VGG16, GoogLeNet i ResNet18, przy czym z uwagi na znaczną złożoność tych sieci ograniczono się do procesu douczania tylko warstw pełnego połączenia. Sieci te wybrano ze względu na ich udokumentowany potencjał w rozwiązywaniu podobnych zadań. Sieć VGG16 została wykorzystana w zadaniu rozpoznawania choroby Parkinsona na podstawie próbek pisma między innymi w pracach [80, 82]. Natomiast w publikacjach [105, 106, 107] sieć GoogLeNet oraz sieci z rodziny ResNet, wykorzystano w rozpoznawaniu choroby Parkinsona na podstawie odręcznych rysunków, nie będących pismem odręcznym. Wykorzystane architektury różnią się między sobą zarówno pod względem budowy jak i liczby parametrów podlegających modyfikacji w procesie uczenia. Porównanie najważniejszych parametrów poszczególnych sieci zamieszczono w tabeli 10.3.

Sieć VGG16 składa się z 16 warstw, w tym 13 warstw konwoucyjnych i 3 warstw w pełni połączonych. Podobnie jak AlexNet ma budowę szeregową, jednak w przeciwieństwie do niej

charakteryzuje się małymi filtrami o rozmiarze 3×3 . Taka strategia pozwala na efektywniejszą ekstrakcję cech, ale znacząco zwiększa liczbę parametrów, co przekłada się na większe wymagania dotyczące mocy obliczeniowej i pamięci. [108]. Sieć GoogLeNet opracowana przez zespół Google, mimo że jest znacznie głębsza od sieci AlexNet, ma zdecydowanie mniejszą liczbę parametrów. Jest to możliwe dzięki zastosowaniu w strukturze tej sieci tzw. modułów *Inception*, które łączą różne rozmiary filtrów w jednej warstwie, umożliwiając równoczesne stosowanie kilku operacji konwolucyjnych w ramach jednego bloku. Takie podejście pozwala na skuteczniejsze uchwycenie cech o różnych rozmiarach i proporcjach, a dodatkowo filtry 1×1 redukują liczbę kanałów, zmniejszając obciążenie obliczeniowe [109]. W strukturze sieci z rodziny ResNet, twórcy zastosowali tzw. połączenia rezydualne, które pozwalają na łatwiejsze trenowanie bardzo głębokich sieci neuronowych, rozwiązując problem zanikania gradientu. Bloki rezydualne składają się z kilku warstw konwolucyjnych, a ich wyjście jest sumowane z wejściem tego bloku, co umożliwia przekazywanie informacji przez sieć bez konieczności uczenia się tożsamościowego odwzorowania. Dzięki takiemu rozwiązaniu sieć może być znacznie głębsza niż wcześniejsze architektury, zachowując efektywność treningu. Na potrzeby rozwiązania zadania niniejszej pracy wybrano najmniejszy model z tej rodziny - ResNet18, składającą się z 18 warstw [110].

Tabela 10.3. Podstawowe informacje dotyczące wykorzystanych sieci.

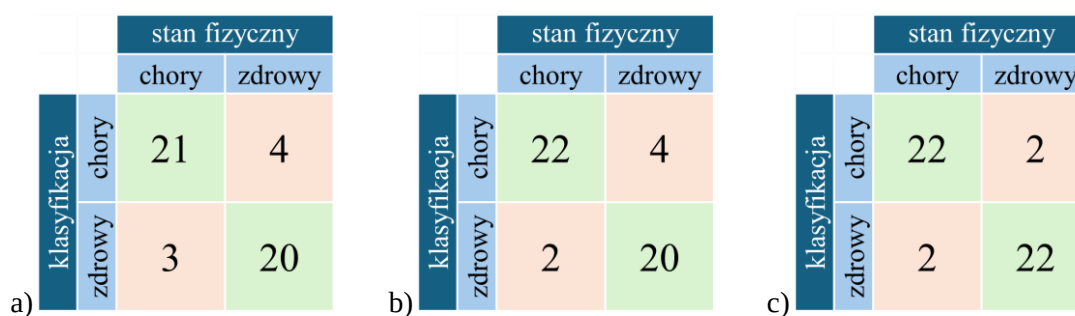
sieć	<i>AlexNet</i>	<i>VGG16</i>	<i>GoogLeNet</i>	<i>ResNet18</i>
liczba warstw	8	16	22	18
liczba parametrów (mln)	61,0	138,0	7,0	11,7
wymiar obrazu wejściowego	227×227	224×224	224×224	224×224
fragment struktury				

W tabeli 10.4. zaprezentowano miary jakości klasyfikacji uzyskane dla różnych sieci konwolucyjnych.

Tabela 10.4. Miary jakości klasyfikacji uzyskane dla różnych sieci konwolucyjnych

sieć	<i>VGG16</i>	<i>GoogLeNet</i>	<i>ResNet18</i>
Dokładność	$0,85 \pm 0,14$	$0,88 \pm 0,12$	$0,92 \pm 0,13$
Precyzja	$0,85 \pm 0,16$	$0,86 \pm 0,15$	$0,93 \pm 0,14$
Czułość	$0,88 \pm 0,17$	$0,92 \pm 0,15$	$0,92 \pm 0,15$
Specyficzność	$0,83 \pm 0,18$	$0,83 \pm 0,18$	$0,92 \pm 0,15$

Niezależnie od wykorzystanej struktury sieci uzyskane dokładności rozpoznania są na wysokim poziomie (między 85 % - 92 %). Najlepszy wynik dokładności rozpoznania uzyskano dla sieci ResNet18, a najniższy dla sieci VGG16. Mimo, że dla wszystkich sieci wynik dokładności wyszedł niższy od uzyskanego dla odpowiednio głęboko dostrojonej struktury sieci AlexNet (93,8 %), to można potwierdzić potencjał każdej z zaproponowanych struktur do rozwiązywania tego typu zadań. Warto zauważyć, że proces dostrajania wag w tym przypadku ograniczono tylko do warstw pełnego połączenia. Zestawiając te wyniki z otrzymanymi dla sieci AlexNet, douczonej również tylko w warstwach pełnego połączenia (patrz tabela 10.2 - *freeze 5*), można zauważyć, że dla zaproponowanych innych modeli sieci uzyskano wyższy wynik dokładności rozpoznania choroby Parkinsona. Zatem również w tym przypadku, odpowiednia modyfikacja głębokości dostrajania wag dla tych modeli, może poprawić wyniki klasyfikacji. Uzyskane macierze błędów dla sieci: VGG16, GoogLeNet i ResNet18 zamieszczono na rysunku 10.6.



Rys. 10.6. Macierze błędów uzyskane z wykorzystaniem sieci: a) VGG16, b) GoogLeNet, c) ResNet18

11. Podsumowanie i wnioski końcowe

W niniejszej rozprawie na potrzeby diagnostyki medycznej podjęto problem klasyfikacji osób zdrowych i osób cierpiących na chorobę Parkinsona na podstawie pisma odręcznego. W badaniach wykorzystano potencjał konwolucyjnych sieci neuronowych, które dzięki charakterystycznej metodzie przetwarzania danych potrafią w sposób automatyczny generować wektor cech na podstawie obrazów wejściowych. Cechą szczególną przeprowadzonych badań było wykorzystanie do uczenia sieci więcej niż jednego obrazu pochodzącego od danej osoby. To podejście pozwoliło zwiększyć liczbę oraz różnorodność danych uczących, celem poprawy działania sieci.

W niniejszej pracy zaprezentowano wyniki trzech podstawowych eksperymentów, których wykonanie pozwoliło potwierdzić słuszność postawionych tez. W pierwszej kolejności przedstawiona została klasyfikacja na podstawie pojedynczego zdania. Badanie miało na celu odniesienie się do wyników z literatury, gdzie podejmowanie decyzji przez sieć następuje na podstawie jednego zdania. Otrzymane porównywalne wyniki stanowiły podstawę dalszych badań. Uzyskano dokładność rozpoznania 60 %. Wykonanie odrębnej analizy dla kolejnych zdań dowiodło, że wydłużenie czasu trwania wysiłku grafomotorycznego pozwala uwidocznić zmiany w piśmie będące konsekwencją występowania choroby Parkinsona. Przy wykorzystaniu obrazów zdania piątego otrzymana dokładność klasyfikacji wynosiła 69 %. Sugerując się tym spostrzeżeniem, wykonano kolejny eksperyment, w którym na wejście sieci zostały podane wszystkie zdania zapisane przez każdą z osób badanych. Zastosowanie tej techniki oraz odpowiedniej integracji decyzji cząstkowych zdecydowanie poprawiło wyniki klasyfikacji (uzyskano dokładność 81 %), co jednoznacznie dowiodło, że **użycie sekwencji zdań daje lepsze wyniki klasyfikacji binarnej w diagnostyce choroby Parkinsona niż użycie zadania pojedynczego**. W kolejnej części badań, przeprowadzono klasyfikację na podstawie pojedynczych słów, podając na wejście sieci obrazy prezentujące wszystkie słowa zapisane w ramach jednej rejestracji. Takie podejście pozwoliło uzyskać dokładność rozpoznania równą 89,6 %, przy zastosowaniu odpowiedniej metody prezentacji danych na obrazie. Rozważania przeprowadzone na podstawie otrzymanych wyników pozwoliły potwierdzić słuszność drugiej z tez, mówiącej, że **możliwa jest dalsza poprawa dokładności klasyfikacji binarnej w diagnostyce choroby Parkinsona na drodze podziału sekwencji zdań na sekwencję słów**.

Dalsze eksperymenty polegające na odpowiednim dostosowaniu struktury sieci AlexNet i dobraniu parametrów uczenia doprowadziły do uzyskania dokładności klasyfikacji równej

93,8 %, co jest wynikiem bardzo obiecującym w porównaniu do wyników uzyskanych przez autorów światowych publikacji w tej dziedzinie. Jednocześnie sieć ta, odpowiednio wykorzystana, zapewniła uzyskanie dokładności rozpoznawania na poziomie nie gorszym niż inne bardziej zaawansowane współczesne struktury sieci konwolucyjnych. W przypadku badań wykorzystujących sieci konwolucyjne tylko autorom [81] udało się uzyskać wyższy wynik rozpoznawania choroby Parkinsona na podstawie próbek pisma odręcznego. Jednak uzyskana przez nich wysoka dokładność 96,58 % może być wynikiem użycia zmodyfikowanych wersji tych samych obrazów do treningu i do testowania co nie jest prawidłową praktyką w przypadku systemów uczących i prowadzi do zbyt optymistycznych wyników. Wynik dokładności rozpoznania uzyskany przez autora niniejszej pracy jest również konkurencyjny dla badań wykorzystujących klasyczną inżynierię cech (w tym również badań autora niniejszej pracy [111]), gdzie wyniki dokładności klasyfikacji osób zdrowych i osób chorych na chorobę Parkinsona na podstawie pisma odręcznego nie przekraczają 92 % [61, 62].

Do głównych osiągnięć autora można zaliczyć:

- Dostosowanie tabletu graficznego do wykonywania badań osób z chorobą Parkinsona.
- Opracowanie oprogramowania do obsługi i rejestracji danych z tabletu.
- Opracowanie algorytmów przetwarzania danych uzyskanych za pośrednictwem tabletu graficznego.
- Opracowanie nowatorskiej koncepcji badań z wykorzystaniem konwolucyjnych sieci neuronowych, mającej na celu poprawę wyników klasyfikacji choroby Parkinsona na podstawie próbek pisma, w stosunku do najlepszych osiągnięć prezentowanych w literaturze.
- Przeprowadzenie serii badań z wykorzystaniem rzeczywistych sygnałów zarejestrowanych w warunkach klinicznych, które wykazały słuszność przedstawionych w rozprawie tez.

Badania przedstawione w pracy pokazują potencjał zastosowania technik głębokiego uczenia w diagnostyce medycznej. Dalsze badania w tym obszarze mogą prowadzić do opracowania jeszcze bardziej skutecznych metod wspomagających wczesne diagnozowanie nie tylko choroby Parkinsona ale również innych chorób neurodegeneracyjnych mogących mieć wpływ na pismo odręczne. Opracowanie nowatorskich rozwiązań opartych na technikach głębokiego uczenia, może umożliwić w przyszłości zautomatyzowanie diagnostyki tych chorób, co pozwoli na skrócenie czasu diagnozy i wcześniejsze rozpoczęcie leczenia. W szczególności interesującym kierunkiem przyszłych badań może być konstrukcja konwolucyjnych modeli regresji zdolnych do przewidywania stopnia nasilenia choroby po jej zdiagnozowaniu. Modele te mogłyby zwracać wartość według skali UPDRS, co stanowiłoby istotne narzędzie w monitorowaniu progresji choroby. Jednakże realizacja takiego podejścia wymagałaby dostępu do większej liczby przypadków odniesienia, aby zapewnić odpowiednią wiarygodność modelu.

Mimo obiecujących wyników, należy podkreślić, że zaproponowany system klasyfikacji może jedynie wspierać proces diagnozowania, lecz nie zastąpi wiedzy i doświadczenia lekarza. Każdy przypadek kliniczny jest unikalny, a decyzja diagnostyczna powinna być podejmowana na podstawie szerokiego spektrum informacji, w tym szczegółowego wywiadu medycznego oraz wyników innych badań. Ponadto, istotnym ograniczeniem metod opisanych w pracy jest konieczność wcześniejszego wykluczenia innych schorzeń, które mogą powodować zaburzenia pisma, lecz nie są związane z chorobą Parkinsona. Wdrożenie takich systemów w praktyce klinicznej wymagałoby również dalszej walidacji oraz oceny ich skuteczności w rzeczywistych warunkach diagnostycznych.

Sama zaprezentowana w pracy metodyka prowadzenia badań z wykorzystaniem sieci konwolucyjnych może być również zastosowana do rozwiązywania innych problemów klasyfikacyjnych opartych na analizie pisma, a nie koniecznie związanych z diagnostyką medyczną np. rozpoznawanie płci czy wieku na podstawie próbek pisma odręcznego.

Bibliografia

- [1] Pang, M., Peng, R., Wang, Y., Zhu, Y., Wang, P., Moussian, B., ... & Ming, D. (2022). Molecular understanding of the translational models and the therapeutic potential natural products of Parkinson's disease. *Biomedicine & Pharmacotherapy*, 155, 113718.
- [2] Feigin, V. L., Nichols, E., Alam, T., Bannick, M. S., Beghi, E., Blake, N., ... & Fischer, F. (2019). Global, regional, and national burden of neurological disorders, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet Neurology*, 18(5), 459-480.
- [3] Kouli, A., Torsney, K. M., & Kuan, W. L. (2018). Parkinson's disease: etiology, neuropathology, and pathogenesis. *Exon Publications*, 3-26.
- [4] Dorsey, E. A., Constantinescu, R., Thompson, J. P., Biglan, K. M., Holloway, R. G., Kieburtz, K., ... & Tanner, C. M. (2007). Projected number of people with Parkinson disease in the most populous nations, 2005 through 2030. *Neurology*, 68(5), 384-386.
- [5] Statistics. Parkinson's Foundation. Pobrane z: <https://www.parkinson.org/understanding-parkinsons/statistics> (dostęp: 17.08.2024).
- [6] Armstrong, M. J., & Okun, M. S. (2020). Diagnosis and treatment of Parkinson disease: a review. *Jama*, 323(6), 548-560.
- [7] Friedman A. (2011). „Etiopatogeneza”, W. J. Sławek, A. Friedman, A. Bogucki, G. Opala (red.), Choroba Parkinsona i inne zaburzenia ruchowe (s. 104-108). Via Medica.
- [8] Massano, J., & Bhatia, K. P. (2012). Clinical approach to Parkinson's disease: features, diagnosis, and principles of management. *Cold Spring Harbor perspectives in medicine*, 2(6).
- [9] Gaenslen, A., Swid, I., Liepelt-Scarfone, I., Godau, J., & Berg, D. (2011). The patients' perception of prodromal symptoms before the initial diagnosis of Parkinson's disease. *Movement Disorders*, 26(4), 653-658.
- [10] Goetz, C. G., Tilley, B. C., Shaftman, S. R., Stebbins, G. T., Fahn, S., Martinez-Martin, P., ... & LaPelle, N. (2008). Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. *Movement disorders: official journal of the Movement Disorder Society*, 23(15), 2129-2170.
- [11] Zach, H., Dirx, M., Bloem, B. R., & Helmich, R. C. (2015). The clinical evaluation of Parkinson's tremor. *Journal of Parkinson's disease*, 5(3), 471-474.
- [12] Beach, T. G., & Adler, C. H. (2018). Importance of low diagnostic accuracy for early Parkinson's disease. *Movement Disorders*, 33(10), 1551-1554.
- [13] Brognara, L., Palumbo, P., Grimm, B., & Palmerini, L. (2019). Assessing gait in Parkinson's disease using wearable motion sensors: a systematic review. *Diseases*, 7(1), 18.
- [14] Balaji, E., Brindha, D., & Balakrishnan, R. (2020). Supervised machine learning based gait classification system for early detection and stage classification of Parkinson's disease. *Applied Soft Computing*, 94.
- [15] Suneel, K., & Kullayamma, I. (2011). ACCELEROMETER SYSTEM FOR TREMOR ANALYSIS. *Global Journal of Engineering and Applied Sciences*, 1(4), 52-55.
- [16] Gunduz, H. (2019). Deep learning-based Parkinson's disease classification using vocal feature sets. *IEEE Access*, 7, 115540-115551.
- [17] Majda-Zdancewicz, E., Dobrowolski, A., Potulska-Chromik, A., Jakubowski, J., Chmieleńska, J., **Bialek, K.**, Nojszewska, M., & Kostera-Pruszczyk, A. (2020). The use of voice processing techniques in the assessment of patients with Parkinson's disease. In *Radioelectronic Systems Conference 2019* (pp. 137-147). SPIE.
- [18] Das, B., Daoudi, K., Klempir, J., & Rusz, J. (2019). Towards disease-specific speech markers for differential diagnosis in parkinsonism. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 5846-5850). IEEE.
- [19] Jakubowski, J., Potulska-Chromik, A., **Bialek, K.**, Nojszewska, M., & Kostera-Pruszczyk, A. (2021). A study on the possible diagnosis of Parkinson's disease on the basis of facial image analysis. *Electronics*, 10(22), 2832.
- [20] Jakubowski, J. M., Potulska-Chromik, A., Chmieleńska, J., Nojszewska, M., & Kostera-Pruszczyk, A. (2023). Application of imaging techniques to objectify the Finger Tapping test used in the diagnosis of Parkinson's disease. *Bulletin of the Polish Academy of Sciences, Technical Sciences*, 71(2).
- [21] Monje, M. H., Domínguez, S., Vera-Olmos, J., Antonini, A., Mestre, T. A., Malpica, N., & Sánchez-Ferro, Á. (2021). Remote evaluation of Parkinson's disease using a conventional webcam and artificial intelligence. *Frontiers in neurology*, 12, 742654.
- [22] Cao, Z., Simon, T., Wei, S. E., & Sheikh, Y. (2017). Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7291-7299).

- [23] Van Galen, G. P. (1991). Handwriting: Issues for a psychomotor theory. *Human movement science*, 10(2-3), 165-191.
- [24] Shukla, A. W., Ounpraseuth, S., Okun, M. S., Gray, V., Schwankhaus, J., & Metzger, W. S. (2012). Micrographia and related deficits in Parkinson's disease: a cross-sectional study. *BMJ open*, 2(3).
- [25] Teulings, H. L., & Stelmach, G. E. (1991). Force amplitude and force duration in parkinsonian handwriting. In *Tutorials in motor neuroscience* (pp. 149-160). Dordrecht: Springer Netherlands.
- [26] Broderick, M. P., Van Gemmert, A. W., Shill, H. A., & Stelmach, G. E. (2009). Hypometria and bradykinesia during drawing movements in individuals with Parkinson's disease. *Experimental brain research*, 197, 223-233.
- [27] Broeder, S., Nackaerts, E., Nieuwboer, A., Smits-Engelsman, B. C., Swinnen, S. P., & Heremans, E. (2014). The effects of dual tasking on handwriting in patients with Parkinson's disease. *Neuroscience*, 263, 193-202.
- [28] Khan, M. A., Ashraf, I., Alhaisoni, M., Damaševičius, R., Scherer, R., Rehman, A., & Bukhari, S. A. C. (2020). Multimodal brain tumor classification using deep learning and robust feature selection: A machine learning application for radiologists. *Diagnostics*, 10(8), 565.
- [29] Khan, M. A., Akram, T., Sharif, M., Kadry, S., & Nam, Y. (2021). Computer Decision Support System for Skin Cancer Localization and Classification. *Computers, Materials & Continua*, 68(1).
- [30] Jojoa Acosta, M. F., Caballero Tovar, L. Y., Garcia-Zapirain, M. B., & Percybrooks, W. S. (2021). Melanoma diagnosis using deep learning techniques on dermatoscopic images. *BMC Medical Imaging*, 21, 1-11.
- [31] Almalki, Y. E., Soomro, T. A., Irfan, M., Alduraibi, S. K., & Ali, A. (2022, April). Computerized analysis of mammogram images for early detection of breast cancer. In *Healthcare* (Vol. 10, No. 5, p. 801). MDPI.
- [32] Kaur, R., GholamHosseini, H., Sinha, R., & Lindén, M. (2022). Melanoma classification using a novel deep convolutional neural network with dermoscopic images. *Sensors*, 22(3), 1134.
- [33] Zham, P., Kumar, D. K., Dabnichki, P., Poosapadi Arjunan, S., & Raghav, S. (2017). Distinguishing different stages of Parkinson's disease using composite index of speed and pen-pressure of sketching a spiral. *Frontiers in neurology*, 8, 435.
- [34] Varalakshmi, P., Priya, B. T., Rithiga, B. A., Bhuvaneaswari, R., & Sundar, R. S. J. (2022). Diagnosis of Parkinson's disease from hand drawing utilizing hybrid models. *Parkinsonism & related disorders*, 105, 24-31.
- [35] Pereira, C. R., Pereira, D. R., Silva, F. A., Masieiro, J. P., Weber, S. A., Hook, C., & Papa, J. P. (2016). A new computer vision-based approach to aid the diagnosis of Parkinson's disease. *Computer methods and programs in biomedicine*, 136, 79-88.
- [36] Taleb, C., Likforman-Sulem, L., Mokbel, C., & Khachab, M. (2023). Detection of Parkinson's disease from handwriting using deep learning: a comparative study. *Evolutionary Intelligence*, 1-12.
- [37] Koziczak, A. (1997). *Metody pomiarowe w badaniach pismoznawczych*. Wydawnictwo Instytutu Ekspertyz Sądowych.
- [38] Kim, E. J., Lee, B. H., Park, K. C., Lee, W. Y., & Na, D. L. (2005). Micrographia on free writing versus copying tasks in idiopathic Parkinson's disease. *Parkinsonism & related disorders*, 11(1), 57-63.
- [39] Miler Jerkovic, V., Kojic, V., Dragasevic Miskovic, N., Djukic, T., Kostic, V. S., & Popovic, M. B. (2019). Analysis of on-surface and in-air movement in handwriting of subjects with Parkinson's disease and atypical parkinsonism. *Biomedical Engineering/Biomedizinische Technik*, 64(2), 187-194.
- [40] Ammour, A., Aouraghe, I., Khaissidi, G., Mrabti, M., Aboulem, G., & Belahsen, F. (2020). A new semi-supervised approach for characterizing the Arabic on-line handwriting of Parkinson's disease patients. *Computer Methods and Programs in Biomedicine*, 183, 104979.
- [41] McLennan, J. E., Nakano, K., Tyler, H. R., & Schwab, R. S. (1972). Micrographia in Parkinson's disease. *Journal of the neurological sciences*, 15(2), 141-152.
- [42] Margolin, D. I., & Wing, A. M. (1983). Agraphia and micrographia: Clinical manifestations of motor programming and performance disorders. *Acta psychologica*, 54(1-3), 263-283.
- [43] Teulings, H. L., & Stelmach, G. E. (1991). Control of stroke size, peak acceleration, and stroke duration in Parkinsonian handwriting. *Human Movement Science*, 10(2-3), 315-334.
- [44] Teulings, H. L., & Stelmach, G. E. (1993). Signal-to-Noise Ratio of Handwriting Size, Force, and Time: Cues to Early Markers of Parkinson's Disease?. In *Sensorimotor impairment in the elderly* (pp. 311-327). Dordrecht: Springer Netherlands.
- [45] Oliveira, R. M., Gurd, J. M., Nixon, P., Marshall, J. C., & Passingham, R. E. (1997). Micrographia in Parkinson's disease: the effect of providing external cues. *Journal of Neurology, Neurosurgery & Psychiatry*, 63(4), 429-433.
- [46] Van Gemmert, A. W. A., Teulings, H. L., Contreras-Vidal, J. L., & Stelmach, G. E. (1999). Parkinsons disease and the control of size and speed in handwriting. *Neuropsychologia*, 37(6), 685-694.

- [47] Van Gemmert, A. W., Teulings, H. L., & Stelmach, G. E. (2001). Parkinsonian patients reduce their stroke size with increased processing demands. *Brain and cognition*, 47(3), 504-512.
- [48] Teulings, H., Contreras-Vidal, J., Stelmach, G., & Adler, C. (2002). Adaptation of handwriting size under distorted visual feedback in patients with Parkinson's disease and elderly and young controls. *Journal of Neurology, Neurosurgery, and Psychiatry*, 72(3), 315.
- [49] Contreras-Vidal, J. L., Teulings, H. L., Stelmach, G. E., & Adler, C. H. (2002). Adaptation to changes in vertical display gain during handwriting in Parkinson's disease patients, elderly and young controls. *Parkinsonism & Related Disorders*, 9(2), 77-84.
- [50] Van Gemmert, A. W. A., Adler, C., & Stelmach, G. (2003). Parkinson's disease patients undershoot target size in handwriting and similar tasks. *Journal of Neurology, Neurosurgery, and Psychiatry*, 74(11), 1502.
- [51] Ünlü, A., Brause, R., & Krakow, K. (2006). Handwriting analysis for diagnosis and prognosis of parkinson's disease. In *Biological and Medical Data Analysis: 7th International Symposium, ISBMDA 2006, Thessaloniki, Greece, December 7-8, 2006. Proceedings 7* (pp. 441-450). Springer Berlin Heidelberg.
- [52] Ponsen, M. M., Daffertshofer, A., Wolters, E. C., Beek, P. J., & Berendse, H. W. (2008). Impairment of complex upper limb motor function in de novo Parkinson's disease. *Parkinsonism & related disorders*, 14(3), 199-204.
- [53] Rosenblum, S., Samuel, M., Zlotnik, S., Erikh, I., & Schlesinger, I. (2013). Handwriting as an objective tool for Parkinson's disease diagnosis. *Journal of neurology*, 260, 2357-2361.
- [54] Smits, E. J., Tolonen, A. J., Cluitmans, L., Van Gils, M., Conway, B. A., Zietsma, R. C., ... & Maurits, N. M. (2014). Standardized handwriting to assess bradykinesia, micrographia and tremor in Parkinson's disease. *PloS one*, 9(5), e97614.
- [55] Zham, P., Raghav, S., Kempster, P., Poosapadi Arjunan, S., Wong, K., Nagao, K. J., & Kumar, D. K. (2019). A kinematic study of progressive micrographia in Parkinson's disease. *Frontiers in neurology*, 10, 403.
- [56] Senatore, R., Marcelli, A., De Micco, R., Tessitore, A., & Teulings, H. L. (2022). Distinctive Handwriting Signs in Early Parkinson's Disease. *Applied Sciences*, 12(23), 12338.
- [57] Drotár, P., Mekyska, J., Rektorová, I., Masarová, L., Smékal, Z., & Faundez-Zanuy, M. (2014). Analysis of in-air movement in handwriting: A novel marker for Parkinson's disease. *Computer methods and programs in biomedicine*, 117(3), 405-411.
- [58] Drotár, P., Mekyska, J., Rektorová, I., Masarová, L., Smékal, Z., & Faundez-Zanuy, M. (2014). Decision support framework for Parkinson's disease based on novel handwriting markers. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 23(3), 508-516.
- [59] Drotár, P., Mekyska, J., Rektorová, I., Masarová, L., Smékal, Z., & Faundez-Zanuy, M. (2016). Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson's disease. *Artificial intelligence in Medicine*, 67, 39-46.
- [60] Impedovo, D., Pirlo, G., & Vessio, G. (2018). Dynamic handwriting analysis for supporting earlier Parkinson's disease diagnosis. *Information*, 9(10), 247.
- [61] Angelillo, M. T., Impedovo, D., Pirlo, G., & Vessio, G. (2019). Performance-driven handwriting task selection for Parkinson's disease classification. In *AI* IA 2019—Advances in Artificial Intelligence: XVIIIth International Conference of the Italian Association for Artificial Intelligence, Rende, Italy, November 19–22, 2019, Proceedings 18* (pp. 281-293). Springer International Publishing.
- [62] Castrillón, R., Acien, A., Orozco-Arroyave, J. R., Morales, A., Vargas, J. F., Vera-Rodríguez, R., ... & Villegas, A. (2019). Characterization of the Handwriting Skills as a Biomarker for Parkinson's Disease. In *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)* (pp. 1-5). IEEE.
- [63] Rios-Urrego, C. D., Vásquez-Correa, J. C., Vargas-Bonilla, J. F., Nöth, E., Lopera, F., & Orozco-Arroyave, J. R. (2019). Analysis and evaluation of handwriting in patients with Parkinson's disease using kinematic, geometrical, and non-linear features. *Computer methods and programs in biomedicine*, 173, 43-52.
- [64] Gupta, U., Bansal, H., & Joshi, D. (2020). An improved sex-specific and age-dependent classification model for Parkinson's diagnosis using handwriting measurement. *Computer methods and programs in biomedicine*, 189, 105305.
- [65] Nolzaco-Flores, J. A., Faundez-Zanuy, M., De La Cueva, V. M., & Mekyska, J. (2021). Exploiting spectral and cepstral handwriting features on diagnosing Parkinson's disease. *IEEE Access*, 9, 141599-141610.
- [66] Parziale, A., Senatore, R., Della Cioppa, A., & Marcelli, A. (2021). Cartesian genetic programming for diagnosis of Parkinson disease through handwriting analysis: Performance vs. interpretability issues. *Artificial Intelligence in Medicine*, 111, 101984.
- [67] Deharab, E. D., & Ghaderyan, P. (2022). Graphical representation and variability quantification of handwriting signals: New tools for Parkinson's disease detection. *Biocybernetics and Biomedical Engineering*, 42(1), 158-172.

- [68] Drotár, P., Mekyska, J., Smékal, Z., Rektorová, I., Masarová, L., & Faundez-Zanuy, M. (2015, May). Contribution of different handwriting modalities to differential diagnosis of Parkinson's disease. In *2015 IEEE international symposium on medical measurements and applications (MeMeA) proceedings* (pp. 344-348). IEEE.
- [69] Mergl, R., Tigges, P., Schröter, A., Möller, H. J., & Hegerl, U. (1999). Digitized analysis of handwriting and drawing movements in healthy subjects: methods, results and perspectives. *Journal of neuroscience methods*, 90(2), 157-169.
- [70] Thomas, M., Lenka, A., & Kumar Pal, P. (2017). Handwriting analysis in Parkinson's disease: current status and future directions. *Movement disorders clinical practice*, 4(6), 806-818.
- [71] Ngo, Q. C., McConnell, N., Motin, M. A., Polus, B., Bhattacharya, A., Raghav, S., & Kumar, D. K. (2024). NeuroDiag: Software for Automated Diagnosis of Parkinson's Disease Using Handwriting. *IEEE Journal of Translational Engineering in Health and Medicine*.
- [72] Van Gemmert, A. W., Teulings, H. L., & Stelmach, G. E. (1998). The influence of mental and motor load on handwriting movements in Parkinsonian patients. *Acta psychologica*, 100(1-2), 161-175.
- [73] Nackaerts, E., Broeder, S., Pereira, M. P., Swinnen, S. P., Vandenberghe, W., Nieuwboer, A., & Heremans, E. (2017). Handwriting training in Parkinson's disease: A trade-off between size, speed and fluency. *PLoS One*, 12(12), e0190223.
- [74] Poluha, P., Teulings, H. L., & Brookshire, R. (1998). Handwriting and speech changes across the levodopa cycle in Parkinson's disease. *Acta psychologica*, 100(1-2), 71-84.
- [75] Tucha, O., Mecklinger, L., Thome, J., Reiter, A., Alders, G. L., Sartor, H., ... & Lange, K. W. (2006). Kinematic analysis of dopaminergic effects on skilled handwriting movements in Parkinson's disease. *Journal of neural transmission*, 113, 609-623.
- [76] Ali, S., Hashmi, A., Hamza, A., Hayat, U., & Younis, H. (2023). Dynamic and static handwriting assessment in Parkinson's disease: a synergistic approach with C-Bi-GRU and VGG19. *Journal of Computing Theories and Applications*, 1(2), 151-162.
- [77] Kumar, K., & Ghosh, R. (2024). Parkinson's disease diagnosis using recurrent neural network based deep learning model by analyzing online handwriting. *Multimedia Tools and Applications*, 83(4), 11687-11715.
- [78] Osowski, S. (2018). Głębokie sieci neuronowe i ich zastosowania w eksploracji danych. *Przegląd Telekomunikacyjny*, 5, 112-121.
- [79] Moetesum, M., Siddiqi, I., Vincent, N., & Cloppet, F. (2019). Assessing visual attributes of handwriting for prediction of neurological disorders—A case study on Parkinson's disease. *Pattern Recognition Letters*, 121, 19-27.
- [80] Diaz, M., Ferrer, M. A., Impedovo, D., Pirlo, G., & Vessio, G. (2019). Dynamically enhanced static handwriting representation for Parkinson's disease detection. *Pattern Recognition Letters*, 128, 204-210.
- [81] Naseer, A., Rani, M., Naz, S., Razzak, M. I., Imran, M., & Xu, G. (2020). Refining Parkinson's neurological disorder identification through deep transfer learning. *Neural Computing and Applications*, 32, 839-854.
- [82] Gazda, M., Hireš, M., & Drotár, P. (2021). Multiple-fine-tuned convolutional neural networks for Parkinson's disease diagnosis from offline handwriting. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 52(1), 78-89.
- [83] Galaz, Z., Drotar, P., Mekyska, J., Gazda, M., Mucha, J., Zvoncak, V., ... & Rektorova, I. (2022). Comparison of CNN-Learned vs. Handcrafted Features for Detection of Parkinson's Disease Dysgraphia in a Multilingual Dataset. *Frontiers in Neuroinformatics*, 16, 877139.
- [84] Osowski, S. (2020). *Sieci neuronowe do przetwarzania informacji* (wyd. 4). Oficyna Wydawnicza Politechniki Warszawskiej.
- [85] Goodfellow, I., Bengio, Y., & Courville, A. (2018). *Deep Learning Systemy uczące się*. Wydawnictwo Naukowe PWN SA.
- [86] Tabor, J., Śmieja, M., Struski, Ł., Spurek, P., & Wołczyk, M. (2022). *Głębokie uczenie wprowadzenie*. Helion SA.
- [87] Krzywicki, T. (2022). *Systemy uczące się. Sztuczne sieci neuronowe i uczenie głębokie*. Wydawnictwo Uniwersytetu Warmińsko-Mazurskiego w Olsztynie.
- [88] Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of big data*, 6(1), 1-48.
- [89] Yang, S., Xiao, W., Zhang, M., Guo, S., Zhao, J., & Shen, F. (2022). Image data augmentation for deep learning: A survey. *arXiv preprint arXiv:2204.08610*.
- [90] Vrbančič, G., & Podgorelec, V. (2020). Transfer learning with adaptive fine-tuning. *IEEE Access*, 8, 196197-196211.
- [91] Nogueira, K., Penatti, O. A., & Dos Santos, J. A. (2017). Towards better exploiting convolutional neural networks for remote sensing scene classification. *Pattern Recognition*, 61, 539-556.

- [92] Chmieleńska, J., **Bialek, K.**, Potulska-Chromik, A., Jakubowski, J., Majda-Zdancewicz, E., Nojszewska, M., ... & Dobrowolski, A. (2020, February). Multimodal data acquisition set for objective assessment of Parkinson's disease. In *Radioelectronic Systems Conference 2019* (pp. 115-124). SPIE.
- [93] Specyfikacja techniczna tabletu WACOM Intuos Pro PTH-860.
- [94] Wacom Intuos Pro: creative pen tablet. Pobrano z: <https://www.wacom.com/en-us/products/pen-tablets/wacom-intuos-pro#Gallery> (dostęp: 21.02.2024).
- [95] Osowski, S. (2013). *Metody i narzędzia eksploracji danych* (wyd. 1). Wydawnictwo BTC.
- [96] Wong, T. T. (2015). Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern recognition*, 48(9), 2839-2846.
- [97] Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling* (Vol. 26, p. 13). New York: Springer.
- [98] Parikh, R., Mathai, A., Parikh, S., Sekhar, G. C., & Thomas, R. (2008). Understanding and using sensitivity, specificity and predictive values. *Indian journal of ophthalmology*, 56(1), 45.
- [99] Fawcett, T. (2006). An introduction to ROC analysis. *Pattern recognition letters*, 27(8), 861-874.
- [100] Cho, M. H., & Hong, C. S. (2015). Two optimal threshold criteria for ROC analysis. *Journal of the Korean Data and Information Science Society*, 26(1), 255-260.
- [101] Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, 3, 32-35.
- [102] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 1097-1105.
- [103] Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks?. *Advances in neural information processing systems*, 27.
- [104] Lee, Y., Chen, A. S., Tajwar, F., Kumar, A., Yao, H., Liang, P., & Finn, C. (2023). Surgical fine-tuning improves adaptation to distribution shifts. *arXiv preprint arXiv:2210.11466*.
- [105] Kamran, I., Naz, S., Razzak, I., & Imran, M. (2021). Handwriting dynamics assessment using deep neural network for early identification of Parkinson's disease. *Future Generation Computer Systems*, 117, 234-244.
- [106] Talitckii, A., Kovalenko, E., Shcherbak, A., Anikina, A., Bril, E., Zimniakova, O., ... & Somov, A. (2022). Comparative study of wearable sensors, video, and handwriting to detect Parkinson's disease. *IEEE Transactions on Instrumentation and Measurement*, 71, 1-10.
- [107] Varalakshmi, P., Priya, B. T., Rithiga, B. A., Bhuvaneaswari, R., & Sundar, R. S. J. (2022). Diagnosis of Parkinson's disease from hand drawing utilizing hybrid models. *Parkinsonism & related disorders*, 105, 24-31.
- [108] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- [109] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1-9).
- [110] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
- [111] **Bialek, K.**, Potulska-Chromik, A., Jakubowski, J., Nojszewska, M., & Kostera-Pruszczyk, A. (2024). Analysis of Handwriting for Recognition of Parkinson's Disease: Current State and New Study. *Electronics*, 13(19), 3962.
- [112] Developer Documentation: Windows Wintab Documentation. Pobrano z: <https://developer-docs.wacom.com/docs/icbt/windows/wintab/wintab-basics/> (dostęp: 21.02.2024).
- [113] Oficjalna strona internetowa twórców dodatku Psychtoolbox-3: Psychtoolbox-3 - Overview. Pobrano z: <http://www.psychtoolbox.org> (dostęp: 21.02.2024)

ZAŁĄCZNIK 1

Szczegółowe informacje kliniczne i demograficzne na temat uczestników badania

ID	pleć	diagnoza	wiek [lata]	skala UPDRS (część III)	czas trwania choroby (od diagnozy) [lata]
1	kobieta	chory	40	10	5
2	mężczyzna	zdrowy	51	-	-
3	kobieta	zdrowy	26	-	-
4	kobieta	zdrowy	50	-	-
5	kobieta	chory	64	34	12
6	mężczyzna	chory	59	28	9
7	kobieta	zdrowy	31	-	-
8	mężczyzna	zdrowy	57	-	-
9	kobieta	chory	28	12	10
10	kobieta	chory	59	10	1
11	kobieta	chory	60	8	2
12	kobieta	chory	64	24	3
13	mężczyzna	chory	44	26	10
14	mężczyzna	chory	44	12	3
15	mężczyzna	chory	73	33	8
16	kobieta	chory	28	16	5
17	mężczyzna	zdrowy	27	-	-
18	mężczyzna	zdrowy	28	-	-
19	kobieta	zdrowy	54	-	-
20	mężczyzna	zdrowy	59	-	-
21	mężczyzna	zdrowy	74	-	-
22	mężczyzna	zdrowy	57	-	-
23	mężczyzna	zdrowy	58	-	-
24	kobieta	chory	62	26	2
25	kobieta	chory	56	22	3
26	kobieta	chory	71	38	15
27	mężczyzna	chory	79	34	10
28	kobieta	chory	61	21	2
29	kobieta	chory	52	24	8
30	kobieta	chory	48	20	6
31	mężczyzna	chory	79	38	10
32	kobieta	chory	61	22	6
33	mężczyzna	chory	73	34	10
34	mężczyzna	chory	81	42	12
35	kobieta	chory	62	24	4
36	kobieta	chory	84	38	12
37	mężczyzna	zdrowy	69	-	-
38	mężczyzna	zdrowy	75	-	-
39	mężczyzna	zdrowy	76	-	-
40	mężczyzna	zdrowy	76	-	-
41	mężczyzna	zdrowy	84	-	-

Szczegółowe informacje kliniczne i demograficzne na temat uczestników badania (cd.)

ID	pleć	diagnoza	wiek [lata]	skala UPDRS (część III)	czas trwania choroby (od diagnozy) [lata]
42	mężczyzna	zdrowy	62	-	-
43	kobieta	zdrowy	58	-	-
44	kobieta	zdrowy	73	-	-
45	kobieta	zdrowy	71	-	-
46	kobieta	zdrowy	66	-	-
47	kobieta	zdrowy	58	-	-
48	kobieta	zdrowy	65	-	-

ZAŁĄCZNIK 2

Autorskie oprogramowanie do rejestracji pisma

Tablety firmy Wacom korzystają z otwartego interfejsu branżowego Wintab Application Programmers Interface (Wintab API), służącego do pozyskiwania danych z tabletów graficznych. Wintab ma na celu zapewnienie stałej łatwej integracji między zaawansowanymi wejściowymi urządzeniami wskazującymi typu tablet a aplikacjami. Interfejs ten umożliwia prostą i jednocześnie bogatą metodologię łączenia aplikacji zgodnych ze standardem Wintab z urządzeniami obsługującymi ten standard. Centralnym elementem systemu jest moduł WINTAB.DLL, umożliwiający usługi interfejsu aplikacji, z którym łączy się co najmniej jeden sterownik urządzenia. Aplikacje manipulują obiektami oprogramowania zwanymi kontekstami, zawierającymi podstawowe informacje na temat sposobu korzystania z tabletu. Gdy aplikacja zdefiniuje i otworzy kontekst Wintab, wyśle on pakiety danych do tej aplikacji w postaci komunikatów Windows, podczas gdy użytkownik naciska piórem na tablet. Obsługę zdarzeń można następnie skonfigurować w celu przechwycenia tych pakietów i wyodrębnienia z nich pożądanego dla aplikacji danych (takich jak np. Współrzędne X, Y, nacisk przykładany na powierzchnię tabletu) [112]. Wykorzystując możliwości interfejsu Wintab, powstał dodatek do jednego z toolbox'ów programu Matlab – Psychtoolbox 3, WinTabMex, który pozwala na bezpośrednie połączenie z tabletem graficznym z poziomu Matlab [113]. Jednym z jego elementów jest plik źródłowy w języku C WinTabMex.c. Plik wymaga dwóch plików nagłówkowych: WINTAB.H i PKTDEF.H. WINTAB.H zawiera definicje wszystkich funkcji i typów danych. PKTDEF.H zawiera sparametryzowaną definicję struktury danych PACKET, którą można dostosować do własnych potrzeb. W tym przypadku zdecydowano się na następującą definicję pakietu danych:

```
#define PACKETDATA (PK_SERIAL_NUMBER | PK_TIME | PK_X | PK_Y | PK_NORMAL_PRESSURE | PK_ORIENTATION)
```

Zawiera on: numer próbki, czas jej pobrania, współrzędną X, współrzędną Y, wartość nacisku oraz informacje dotyczące orientacji rysika, w skład których wchodzi kąt nachylenia pióra do powierzchni tabletu oraz azymut, czyli kąt zawarty między kierunkiem pióra a osią Y tabletu. Plik WinTabMex.c korzysta z wielu funkcji podstawowych, które obejmują uzyskiwanie informacji o interfejsie i urządzeniu, otwieranie i zamykanie kontekstów oraz pobieranie pakietów. Najważniejsze wykorzystywane funkcje opisano w tabeli 1.

Tabela 1. Opis funkcji do obsługi interfejsu WinTab

funkcja	opis
WTInfo()	Zwraca globalne informacje o interfejsie takie jak: typ tabletu, wymiary fizyczne tabletu, rozdzielczość tabletu, możliwości i typy kursorów.
WTOpen()	Ustawia aktywny kontekst na tablecie. Po pomyślnym zakończeniu tej funkcji aplikacja może zacząć odbierać zdarzenia tabletu.

Tabela 1. Opis funkcji do obsługi interfejsu WinTab (cd.)

funkcja	opis
WTClose()	Zamyka i niszczy obiekt kontekstu tabletu.
WTPacketsGet()	Kopiuje pakiety danych i opróżnia kolejkę zdarzeń, ze starych zdarzeń w ramach przygotowań do gromadzenia nowych danych.
WTEnable()	Włącza lub wyłącza kontekst tabletu.

Pakiet WinTabMex zawarty w Psychtoolbox 3 składa się z 6 funkcji do obsługi tabletu bezpośrednio opierających się o funkcje zapisane w tabeli 2. Są to:

Tabela 2. Opis funkcji WinTabMex

funkcja WinTabMex	opis	wykorzystywane funkcje Wintab
WinTabMex(0,wPtr)	inicjalizacja sterownika, otwarcie dostępu do pozostałych funkcji	
WinTabMex(1)	wyłączenie dostępu Matlaba do sterownika tabletu	WTClose()
WinTabMex(2)	przygotowanie do gromadzenia danych poprzez wyczyszczenie kolejki zdarzeń, ze starych zdarzeń	WTPacketsGet(), WTEnable()
WinTabMex(3)	zatrzymanie gromadzenia danych	WTEnable()
WinTabMex(4)	zmiana rozmiaru kolejki zdarzeń (tutaj niewykorzystywana)	WTQueueSizeSet()
WinTabMex(5)	odczytanie najstarszego pakietu danych i zwrócenie siedmio kolumnowej macierzy, zawierającej dane pochodzące z tabletu	WTPacketsGet()

Są to gotowe funkcje możliwe do wywołania z poziomu Matlaba podczas standardowego generowania programu. Wykorzystując możliwości pakietu WinTabMex napisano autorskie oprogramowanie do obsługi i zbierania danych z tabletu PTH-860 w środowisku Matlab R2018b firmy Mathworks. Opracowany kod przedstawiono w tabeli 3.

Tabela 3. Kod programu obsługującego tablet graficzny

```

%% Obsługa tabletu

% Definicja przycisku - SPACJI do obsługi tabletu
KbName('UnifyKeyNames');
spaceKeyID = KbName('space');

% Inicjalizacja
wPtr = Screen('OpenWindow', 0);
WinTabMex(0, wPtr); %Inicjalizacja sterownika tabletu, załadowanie okna wPtr

% Elementy Psychtoolbox
Priority(1);
ListenChar(2); %Włączenie obsługi przycisków klawiatury
HideCursor; % Ukrywanie kursora

% Wyświetlenie komunikatu
instructions = 'Umieść rysik na tablecie i wciśnij spację, aby rozpocząć nagrywanie';
[instructionsX instructionsY] = centreText(wPtr, instructions, 15);
Screen('DrawText', wPtr, instructions, instructionsX, instructionsY, BlackIndex(wPtr));
Screen(wPtr, 'Flip');

```

Tabela 3. Kod programu obsługującego tablet graficzny (cd.)

```

%Oczekiwanie na wciśnięcie przycisku
[keyIsDown, secs, keyCode] = KbCheck;
while keyCode(spaceKeyID)~=1
    [keyIsDown, secs, keyCode] = KbCheck;
end

% Wyświetlenie komunikatu o zbieraniu danych
record = 'Tablet rejestruje twoje ruchy';
[recordX recordY] = centreText(wPtr, record, 15);
Screen('DrawText', wPtr, record, recordX, recordY, BlackIndex(wPtr));
Screen('DrawDots', wPtr,[100 100], 20, [1 0 0], [], 2); Screen(wPtr, 'Flip');

% Zebranie danych
[daneaktualne]=akwizycjadanych(200, wPtr);

% Zamykanie połączenia
WinTabMex(1); % Wyłączenie sterownika
ListenChar(1); % Wyłączenie obsługi przycisków klawiatury
ShowCursor; % Pokazanie kursora
Screen('CloseAll');

% Odtworzenie obrazu
[obraz]=odtworzenieobrazu(daneaktualne,20);
% Zapis danych do pliku
zapisdanych(daneaktualne);

```

Przyciskiem wybranym do rozpoczęcia i zakończenia procesu rejestracji jest spacja, która została zdefiniowana na początku kodu programu. Aby możliwa była obsługa poprzez klawisz klawiatury należy włączyć obsługę przycisków *ListenChar(2)* po inicjalizacji sterownika *WinTabMex(0, wPtr)* oraz ukryć kursor *HideCursor*. Po uruchomieniu programu na całym ekranie pojawia się okno z komunikatem: „Umieść rysik na tablecie i wciśnij spację, aby rozpocząć nagrywanie”. W momencie wciśnięcia spacji tablet rozpoczyna rejestrację ruchów poprzez wykonywanie funkcji *akwizycjadanych()*, z częstotliwością zadeklarowaną jako parametr tej funkcji, a na ekranie wyświetlony jest komunikat: „Tablet rejestruje twoje ruchy”. Po wykonaniu rysunku i ponownym wciśnięciu spacji, sterownik zostaje wyłączony *WinTabMex(1)*, następuje wyłączenie obsługi przycisków *ListenChar(1)*, kursor ponownie jest widoczny *ShowCursor*, a ekran aplikacji zostaje zamknięty *Screen('CloseAll')*. Po zamknięciu aplikacji następuje odtworzenie rysunku utworzonego na powierzchni tabletu na podstawie zarchiwizowanych danych i ich zapis do pliku pod nazwą: *IDXX_dane.mat*, gdzie *XX* oznacza numer identyfikacyjny pacjenta. Funkcja *akwizycjadanych(częstotliwość próbkowania, udostępniany wątek)* służy do zbierania danych z tabletu. Jednym z argumentów tej funkcji jest częstotliwość próbkowania, wynosząca tutaj 200 Hz, co jest maksymalną częstotliwością próbkowania dla tabletu PTH-860. W tabeli 4 przedstawiono kod funkcji *akwizycjadanych(częstotliwość próbkowania, udostępniany wątek)*.

Tabela 4. Kod źródłowy funkcji do akwizycji danych

```

function [pktData]=akwizycjadanych(samplingRate, wPtr)
deltaT = 1/samplingRate;

pktData = []; % Konfiguracja zmiennej do przechowywania danych

```

Tabela 4. Kod źródłowy funkcji do akwizycji danych (cd.)

```

% Pętla odpowiedzialna za pobieranie danych
try
    WinTabMex(2); % Przygotowanie do zbierania danych
    start = GetSecs;
    while ~KbCheck % Sprawdzenie statusu klawisza
        loopStart = GetSecs;
        while 1
            pkt = WinTabMex(5); % Zbieranie danych z tabletu
            if ~isempty(pkt)
                break
            end
            if GetSecs>(loopStart+deltaT) % Sprawdzenie istnienia punktu
                pkt = zeros(7,1); % Dane reprezentujące pominięty punkt danych
                break;
            end
        end
        pktData = [pktData pkt];
        if GetSecs<(loopStart+deltaT) % Oczekiwanie na zakończenie czasu
            WaitSecs('UntilTime', loopStart+deltaT);
        end
    end
    pktData = pktData'; % Złożenie danych i ich transpozycja
    WinTabMex(3); % Stop/Pauza akwizycji danych.
    Priority(0);
end

```

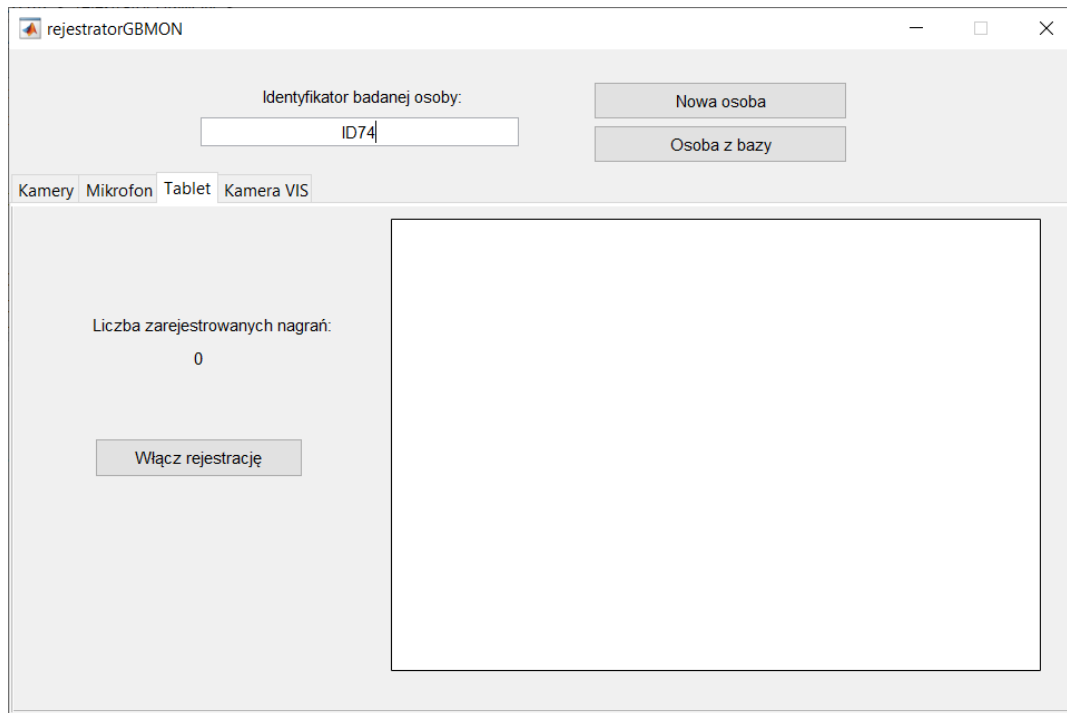
Funkcja ta zwraca macierz wartości *pktData*, w której każdy wiersz reprezentuje pojedynczy pakiet danych pomiarowych, a kolumny poszczególne jego wartości. Zawierają one następujące informacje: numer próbki, czas rejestracji w [ms], położenie pióra wzdłuż osi X oraz Y tabletu, nacisk w jednostkach umownych, kąt nachylenia pióra do powierzchni tabletu w [°] oraz kąt zawarty między osią Y tabletu, a piórem tzw. azymut w [°]. W tabeli 5 zaprezentowano fragment zebranych przykładowych danych.

Tabela 5. Przykładowy zestaw danych rejestrowanych przez wykorzystany tablet graficzny

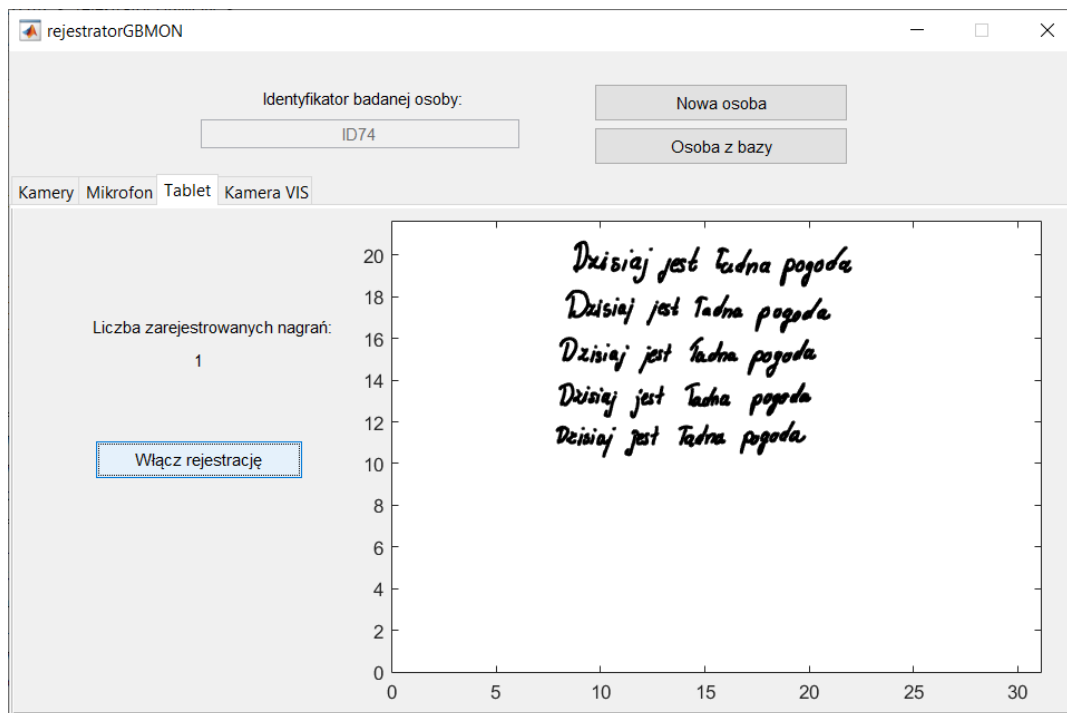
numer próbki	znacznik czasu [ms]	współrzędna x [cm]	współrzędna y [cm]	nacisk [j.u.]	kąt nachylenia [°]	azymut [°]
1	0	2,685	18,612	148	51	83
2	5	2,685	18,616	5700	51	83
3	10	2,685	18,619	6912	51	83
4	15	2,685	18,623	7696	51	81
5	20	2,685	18,624	8952	50	81
6	25	2,686	18,625	10220	50	80
7	30	2,686	18,625	10868	50	80
8	35	2,687	18,626	11265	49	80
9	40	2,687	18,626	11750	50	79

Powyższe funkcje zostały wykorzystane do opracowania finalnego oprogramowania obsługującego stanowisko do pomiaru multimodalnych danych na potrzeby obiektywnej oceny choroby Parkinsona [92]. Okno aplikacji służące do rejestrowania danych w postaci pisma

odrębnego z tabletu graficznego pokazano na rysunku 1 (widoczne są zakładki odpowiedzialne za rejestrację pozostałych modalności).



a)



b)

Rys. 1. Okno programu do rejestracji multimodalnych danych, otwarte na zakładce Tablet: a) przed rozpoczęciem rejestracji, b) po wykonaniu rejestracji