



**UNIwersYTET
MIKOŁAJA KOPERNIKA
W TORUNIU**

Collegium Medicum
im. Ludwika Rydygiera w Bydgoszczy

Bydgoszcz 2024



**UNIWERSYTET
MIKOŁAJA KOPERNIKA
W TORUNIU**
Wydział Lekarski
Collegium Medicum w Bydgoszczy

Adrian Michalski

Ocena przydatności metod uczenia maszynowego w systemach wspierających decyzje diagnostyczne

Rozprawa na stopień doktora nauk medycznych i nauk o zdrowiu

Promotor:

Dr hab. Bogumiła Kupcewicz, prof. UMK

Bydgoszcz 2024

Spis treści

1.	Lista prac wchodzących w skład rozprawy	4
2.	Wykaz skrótów.....	5
3.	Wstęp.....	7
3.1.	Sztuczna inteligencja	8
3.2.	Uczenie maszynowe	10
3.3.	Uczenie głębokie.....	12
3.4.	Systemy wsparcia decyzji diagnostycznych	15
3.5.	Systemy wsparcia decyzji diagnostycznych w ścieżce medycznej pacjenta	16
3.5.1.	Podstawowa opieka zdrowotna.....	18
3.5.2.	Środowisko szpitalne	20
3.5.3.	Laboratorium	22
3.5.4.	Diagnostyka obrazowa	24
3.5.5.	Planowanie i monitorowanie leczenia	25
3.6.	Duże modele językowe w systemach wsparcia decyzji klinicznych	27
3.7.	Wytlumaczalność modeli sztucznej inteligencji	28
3.8.	Systemy wsparcia decyzji diagnostycznych w chorobach rzadkich	29
3.8.1.	Choroba Andersona-Fabry'ego	30
4.	Cel pracy.....	31
5.	Omówienie prac włączonych do cyklu	32
5.1.	Artykuł oryginalny I	32
5.2.	Artykuł oryginalny II	36
6.	Wnioski.....	41
7.	Bibliografia	43
	Streszczenie	49
	Summary	50
	Życiorys zawodowy	51
	Załącznik I	54
	Załącznik II	61
	Załącznik III.....	74

1. Lista prac wchodzących w skład rozprawy

W skład niniejszej rozprawy wchodzi 2 prace oryginalne opublikowane w recenzowanych czasopismach umieszczonych w wykazie czasopism naukowych i recenzowanych materiałów z konferencji międzynarodowych. Sumaryczny Impact Factor publikacji stanowiących rozprawę wynosi 8,414 i 210 punktów MNiE (Ministerstwa Nauki i Edukacji).

Michalski A, Duraj K, Kupcewicz, B. *Leukocyte deep learning classification assessment using Shapley additive explanations algorithm*. Int J Lab Hematol. 2023; 1-6.
<https://doi.org/10.1111/ijlh.14031>

IF = 3,45 MNiSW = 70

Michalski A, Lis K, Stankiewicz J, Kloska S.M, Dudziński M, Muras-Szwedziak K, Nowicki M, Bazan-Socha S, Dąbrowski M.J, Basak G.W. *Supporting the diagnosis of Fabry disease using a natural language processing-based approach*. J. Clin. Med. 2023, 12(10), 3599;
<https://doi.org/10.3390/jcm12103599>

IF = 4,964 MNiSW = 140

2. Wykaz skrótów

AFD – choroba Andersona-Fabr'yego (ang. Anderson-Fabry Disease)

AI – sztuczna inteligencja (ang. Artificial Intelligence)

ALT – aminotransferaza alaninowa (ang. Alanine Transaminase)

ANN – sztuczna sieć neuronowa (ang. Artificial Neural Network)

AR – rzeczywistość rozszerzona (ang. Augmented Reality)

AUC – pole pod krzywą (ang. Area Under Curve)

BERT – (ang. Bidirectional Encoder Representations from Transformers)

CAGR - roczna skumulowana stopa wzrostu (ang. Compound Annual Growth Rate)

CNN – splotowe sieci neuronowe (ang. Convolutional Neural Networks)

CRP – białko c-reaktywne (ang. C-reactive Protein)

CT – tomografia komputerowa (ang. Computed Tomography)

CV – wizja komputerowa (ang. Computer Vision)

DDSS – system wsparcia decyzji diagnostycznych (ang. Diagnostic Decision Support System)

DL – uczenie głębokie (ang. Deep Learning)

FBP – (ang. Filtered Backpropagation)

FDA - agencja żywności i leków (ang. Food and Drug Administration)

GAN – generatywne sieci antagonistyczne (ang. Generative Adversarial Network)

Gb3 – globotriaosylceramid (ang. Globotriaosylceramide)

GD – choroba Gaucher (ang. Gaucher Disease)

GPT – (ang. Generative Pre-trained Transformer)

IR - rekonstrukcja iteracyjna (ang. Iterative Reconstruction)

KNN – algorytm k najbliższych sąsiadów (ang. K-Nearest Neighbors)

LIME – (ang. Local Interpretable Model-Agnostic Explanations)

LLM – duże modele językowe (ang. Large Language Models)

LSTM – (ang. Long Short-Term Memory)

ML – uczenie maszynowe (ang. Machine Learning)

MR - rezonans magnetyczny (ang. Magnetic Resonance)

NIH - narodowego instytutu zdrowia (ang. National Institute of Health)

NLP – przetwarzanie języka naturalnego (ang. Natural Language Processing)

PID – pierwotne niedobory odporności (ang. Primary Immunodeficiencies)

RNN – rekurencyjne sieci neuronowe (ang. Recurrent Neural Networks)

RTG – rentgenogram

SHAP – (ang. SHapley Additive exPlanations)

SVM – maszyna wektorów nośnych (ang. Support Vector Machine)

TMA – mikroangiopatia zakrzepowa (ang. Thrombotic Microangiopathy)

TTP - zakrzepowa plamica małopłytkowa (ang. Thrombotic Thrombocytopenic Purpura)

USG – ultrasonografia (ang. Ultrasonography)

xAI – wytłumaczalna sztuczna inteligencja (ang. Explainable Artificial Intelligence)

3. Wstęp

W 2018 roku, w momencie rozpoczęcia moich badań nad systemami wsparcia decyzji diagnostycznych (diagnostic decision support systems, DDSS) wykorzystującymi metody uczenia maszynowego (machine learning, ML), wartość rynku sztucznej inteligencji (artificial intelligence, AI) w ochronie zdrowia wynosiła 2.5 miliarda dolarów. Przewidywano, że roczna skumulowana stopa wzrostu (compound annual growth rate, CAGR) w 2018-2025 wyniesie 41.5%, co daje na rok 2024 kwotę 20.07 miliarda dolarów [1]. W rzeczywistości wartość rynku AI w ochronie zdrowia wynosi w 2024 roku 32.3 miliarda dolarów i przewidywany jest jej dalszy wzrost do 208.2 miliarda dolarów w 2030 roku [2]. Taka dynamika wpisuje się w ogólny trend rozwoju sztucznej inteligencji i jest silnym argumentem przemawiającym za wykorzystaniem AI w praktyce diagnozowania, personalizacją terapii i zarządzaniem danymi klinicznymi.

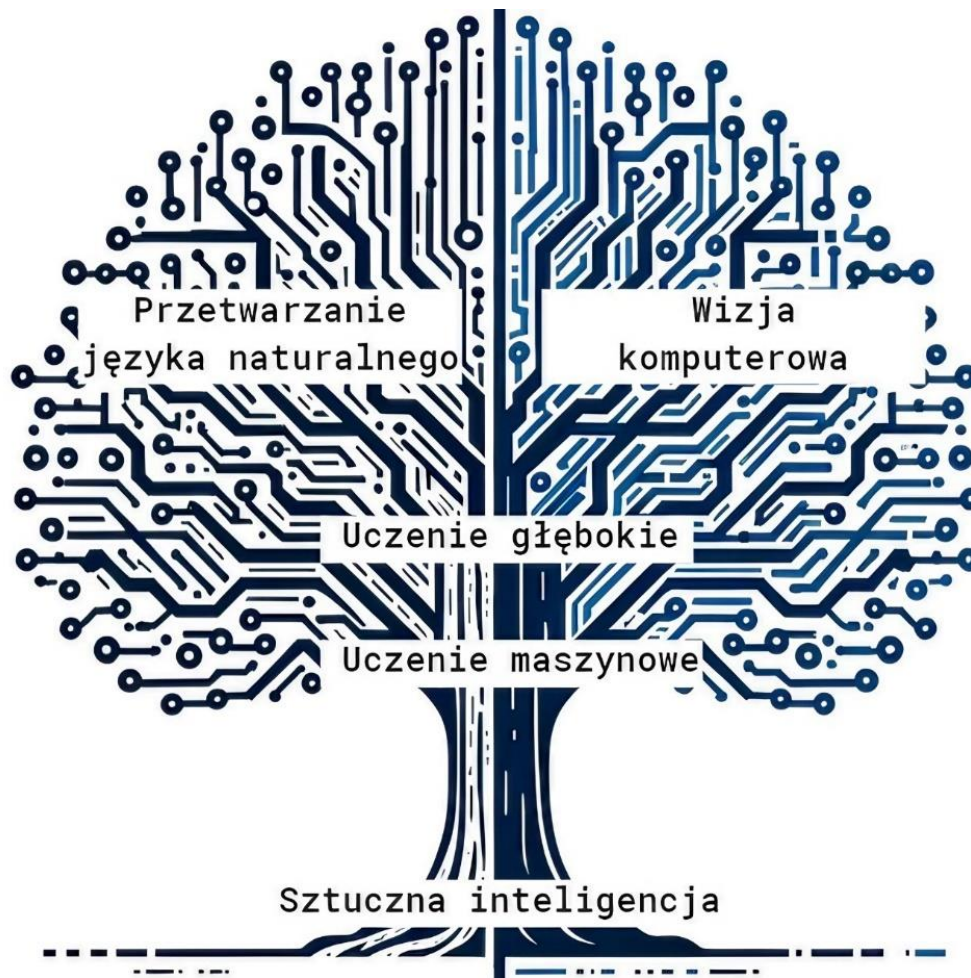
W niniejszej pracy przedstawię zastosowanie uczenia maszynowego w systemach wsparcia decyzji diagnostycznych, wskazując miejsca w ścieżce pacjenta, w których sztuczna inteligencja wspiera lub mogłaby wesprzeć proces diagnostyczny, zwiększając dokładność diagnoz, skracając czas ich stawiania oraz ułatwiając identyfikację rzadkich lub złożonych chorób [3]. W ramach pracy przybliżę rolę uczenia maszynowego w DDSS poprzez analizę jego zastosowania w wytłumaczalności modeli wizji komputerowej do analizy obrazów mikroskopowych rozmazów krwi oraz w przetwarzaniu języka naturalnego do ekstrakcji informacji z historii medycznej pacjenta i podejmowania decyzji diagnostycznych. Przedmiotem szczególnego zainteresowania w niniejszej pracy jest zastosowanie uczenia maszynowego w diagnostyce chorób rzadkich. Ze względu na niską częstość występowania i zróżnicowaną symptomatologię, choroby rzadkie często stanowią wyjątkowe wyzwanie w zakresie diagnostyki i leczenia [4, 5].

3.1. Sztuczna inteligencja

W obszarze sztucznej inteligencji mieszczą się wzajemnie powiązane ze sobą dziedziny takie jak uczenie maszynowe, uczenie głębokie (deep learning, DL), przetwarzanie języka naturalnego (natural language processing, NLP) i wizja komputerowa (computer vision, CV) (Rycina 1). Uczenie maszynowe stanowi dziedzinę sztucznej inteligencji (artificial intelligence, AI) skoncentrowaną na rozwoju algorytmów, które umożliwiają komputerom autonomiczne uczenie się, przewidywanie i podejmowanie decyzji. Systemy uczenia maszynowego wykazują zdolność do autonomicznej poprawy wydajności w czasie, bez konieczności ingerencji programisty [6]. Uczenie głębokie to poddziedzina uczenia maszynowego, która czerpie inspirację z biologii, wykorzystując sztuczne sieci neuronowe (artificial neural network, ANN) do imitowania sposobu, w jaki ludzki mózg przetwarza informacje. Sieci neuronowe zbudowane są z warstw połączonych neuronów, nazywanych również węzłami, które imitują działanie biologicznych neuronów. Te sztuczne sieci są w stanie uczyć się złożonych wzorców i reprezentacji danych, co czyni je niezwykle potężnym narzędziem w wielu dziedzinach nauki [7].

NLP to gałąź AI, która skupia się na zrozumieniu i interpretacji języka ludzkiego. Zajmuje się zadaniami, takimi jak systemy dialogowe, analiza tekstu, analiza sentymentu i tłumaczenie dokumentów. NLP wykorzystuje techniki uczenia maszynowego i uczenia głębokiego do modelowania złożoności ludzkiego języka i tworzenia bardziej inteligentnych, podobnych do ludzkich interakcji między komputerami i użytkownikami [8].

CV to obszar AI, który umożliwia komputerom interpretację i zrozumienie informacji wizualnych. Obejmuje ona takie zadania, jak wykrywanie obiektów, rozpoznawanie wzorców, segmentacja i generowanie obrazów [9].



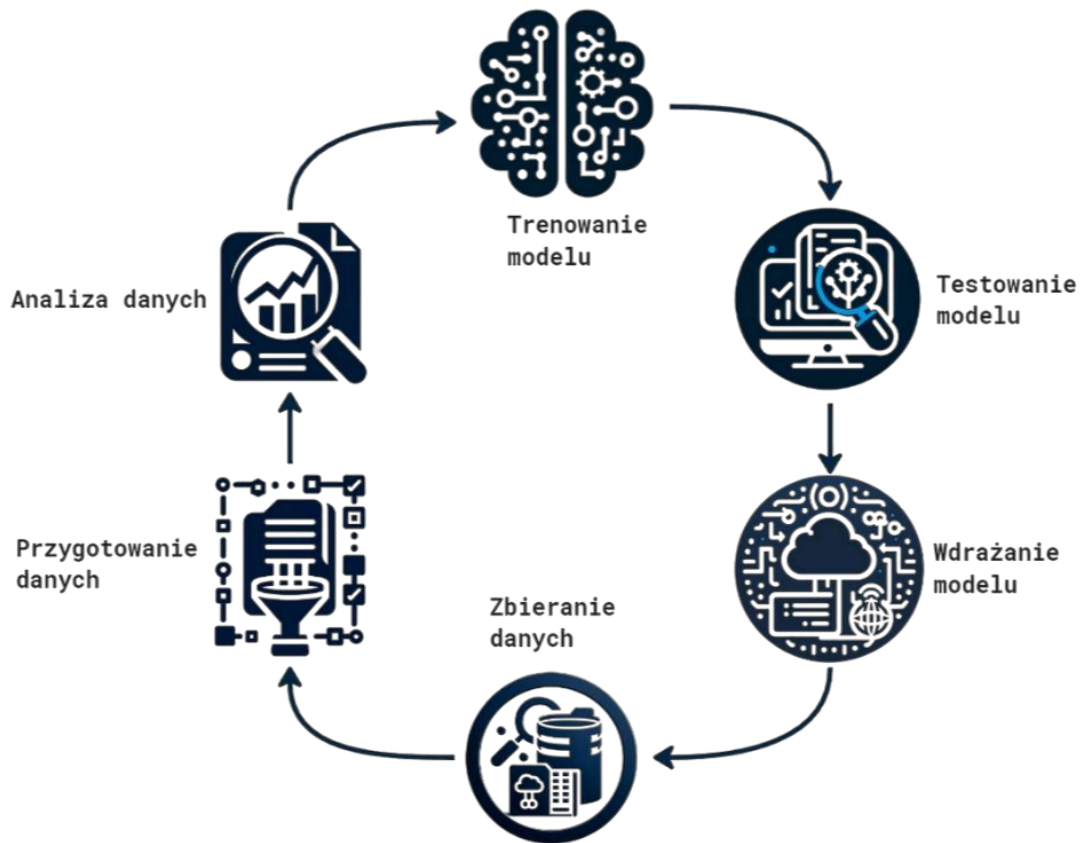
Rycina 1. Drzewo zależności w obszarze sztucznej inteligencji

3.2. Uczenie maszynowe

Uczenie maszynowe koncentruje się na rozwoju algorytmów i technik umożliwiających komputerom uczenie się, prognozowanie i podejmowanie decyzji na podstawie danych. Podstawową ideą ML jest tworzenie modeli na podstawie danych. Rozwój tych modeli obejmuje zazwyczaj następujące etapy: gromadzenie i wstępne przetwarzanie danych, szkolenie modelu, ocenę modelu i jego wdrożenie (Rycina 2). Gromadzenie i wstępne przetwarzanie danych polega na zbieraniu surowych danych, oczyszczaniu ich i przekształcaniu w odpowiedni format do szkolenia modelu. Szkolenie modelu sztucznej inteligencji to proces, w którym uczy się on rozpoznawać wzorce w danych np. klasyfikacja etiologii zapalenia płuc pacjenta na podstawie podstawowych parametrów laboratoryjnych, zdjęć RTG klatki piersiowej i objawów pacjenta. Ewaluacja modelu polega na ocenie wydajności wytrenowanego modelu, natomiast wdrożenie modelu to proces integracji wytrenowanego modelu z aplikacją [9]

Istnieje kilka czynników, które przyczyniają się do skuteczności modeli uczenia maszynowego, w tym jakość i ilość danych treningowych, wybór algorytmu uczenia maszynowego oraz proces inżynierii cech. Wysokiej jakości, zróżnicowane i reprezentatywne dane treningowe są niezbędne, aby modele uczenia maszynowego dobrze generalizowały i dostarczały dokładnych wyników. Wybór algorytmu uczenia maszynowego zależy od konkretnego problemu, który jest rozwiązywany i rodzaju danych. Inżynieria cech to proces wyboru, przekształcania lub tworzenia nowych cech z surowych danych.

Uczenie maszynowe można podzielić na trzy główne kategorie w zależności od zastosowanej techniki uczenia: uczenie nadzorowane, uczenie bez nadzoru i uczenie ze wzmocnieniem. Uczenie nadzorowane wykorzystuje dane z etykietami, gdzie pary wejście-wyjście są znane [10]. Typowe zadania uczenia nadzorowanego obejmują regresję i klasyfikację. Z kolei uczenie nienadzorowane polega na znajdowaniu wzorców i zależności w danych nieoznakowanych, gdzie pary wejście-wyjście nie są znane. Typowymi zadaniami uczenia bez nadzoru jest redukcja wymiarowości i klasteryzacja [11]. Uczenie wzmocniające polega na tym, że model uczy się podejmować decyzje na podstawie interakcji ze środowiskiem i otrzymuje informacje zwrotne w postaci nagród lub kar [12].



Rycina 2. Cykl życia modelu uczenia maszynowego.

3.3. Uczenie głębokie

Uczenie głębokie skupia się na rozwoju i zastosowaniu sztucznych sieci neuronowych z wieloma warstwami. Te głębokie architektury mogą przetwarzać duże ilości danych i uczyć się złożonych wzorców. Uczenie głębokie zyskało znaczne zainteresowanie w ostatnich latach ze względu na zdolność do osiągnięcia wysokiej skuteczności i wydajności w wielu dziedzinach takich jak przetwarzaniu języka naturalnego i rozpoznawanie obrazów i mowy [7].

Kluczową koncepcją uczenia głębokiego jest wykorzystanie hierarchicznych warstw sztucznych neuronów. Każda warstwa w głębokiej sieci neuronowej składa się z wielu połączonych ze sobą neuronów, które przetwarzają dane wejściowe i przekształcają je w bardziej abstrakcyjną reprezentację. Poprzez układanie wielu warstw, modele uczenia głębokiego mogą uczyć się coraz bardziej złożonych i abstrakcyjnych cech, co pozwala im uchwycić zawite wzorce w danych. Istnieje kilka rodzajów architektur głębokiego uczenia, w tym [13]:

Splotowe sieci neuronowe (convolutional neural networks, CNN). CNN, zaprojektowane specjalnie do przetwarzania i analizy obrazów, odgrywają kluczową rolę w aplikacjach wizji komputerowej. Ich unikalna struktura opiera się na wykorzystaniu warstw splotowych, które zawierają filtry przetwarzające dane wejściowe. Te filtry skutecznie identyfikują i uczą się różnych wzorców wizualnych, takich jak krawędzie, kształty czy tekstury. W CNN znajdują się również warstwy łączące (pooling layers), które redukują wymiary przestrzenne danych. Dzięki temu procesowi, CNN są w stanie lepiej radzić sobie ze zmianami skali i pozycji obiektów na obrazie, co przekłada się na wyższą odporność i efektywność modeli w różnych warunkach. CNN mają zastosowania w szerokim zakresie zadań, od rozpoznawania twarzy po analizę obrazów medycznych [14].

Rekurencyjne sieci neuronowe (recurrent neural network, RNN). RNN zostały zaprojektowane do obsługi danych sekwencyjnych i są szczególnie skuteczne w zadaniach przetwarzania języka naturalnego i analizy szeregów czasowych. Zawierają one pętle sprzężenia zwrotnego, które pozwalają sieci na utrzymanie ukrytego stanu, co umożliwia uchwycenie czasowych zależności w danych [15].

Sieci LSTM (long short-term memory). LSTM są szczególnym rodzajem sieci RNN, które rozwiązują problem znikającego gradientu, występującego, gdy gradienty w sieci stają się zbyt małe, aby skutecznie aktualizować wagi cech. Sieci LSTM wykorzystują wyspecjalizowane komórki pamięci i mechanizmy bramkowania, które pozwalają im zachowywać i uczyć się długoterminowych zależności w danych, co czyni je odpowiednimi do zadań obejmujących analizę długich sekwencji np. bloków tekstu [16].

Transformery. Transformery są architekturą uczenia głębokiego, którą w ostatnich latach szeroko stosowano w przetwarzaniu języka naturalnego. Zostały zaprojektowane do równoległego przetwarzania sekwencji, pokonując ograniczenia RNN i LSTM w zakresie szybkości obliczeń i skalowalności. Transformery wykorzystują mechanizm uwagi (self-attention mechanism), który pozwala im ważyć znaczenie elementów w obrębie sekwencji, skuteczniej wychwytyjąc kontekst oraz zależności między odległymi informacjami [17].

Duże modele językowe (large language models, LLM). Na szczególną uwagę zasługują obecnie rozwijane LLM takie jak GPT (generative pre-trained transformer) czy BERT (bidirectional encoder representations from transformers). LLM są zbudowane na architekturze transformerów i osiągnęły najlepszą dotychczas wydajność w szerokim zakresie zadań NLP. Modele te są wstępnie trenowane na ogromnych ilościach danych tekstowych, co pozwala im uczyć się skomplikowanych wzorców i związków w języku. Dostrojenie tych modeli do konkretnych zadań pozwala im na uogólnienie wiedzy i zapewnienie podobnego do ludzkiego zrozumienia i wydajności. LLM wywarły znaczny wpływ na dziedzinę AI, demonstrując moc uczenia transferowego oraz potencjał wykorzystania na dużą skalę w wielu zastosowaniach [18].

W niektórych dziedzinach istotny jest nie tylko wynik działania modelu, ale i odpowiedź dlaczego model podjął właśnie taką decyzję. W uczeniu głębokim to pytanie jest szczególnie trudne, a w ochronie zdrowia odpowiedź na nie szczególnie ważna. Za każdym razem kiedy model sklasyfikuje pacjenta jako chorego lub zdrowego, powinniśmy otrzymać informację na jakiej podstawie model podjął tę decyzję. W szczególności, w przypadku diagnozowania chorób rzadkich i wielochorobowości, zrozumienie procesu decyzji modelu może pomóc lekarzom w lepszym rozumieniu patologii pacjenta i podejmowaniu szybszych, skutecznych decyzji terapeutycznych. W konsekwencji, w ochronie

zdrowia kluczowe jest nie tylko zastosowanie modeli uczenia głębokiego, ale również rozwijanie metod wyjaśniania ich decyzji, aby zapewnić pełniejsze zrozumienie procesu diagnozowania i leczenia chorób [19].

3.4. Systemy wsparcia decyzji diagnostycznych

Systemy wsparcia decyzji diagnostycznych to zautomatyzowane systemy zaprojektowane w celu wspomaganie pracowników ochrony zdrowia w diagnozowaniu chorób. Korzystając z różnorodnych metod analizy danych, w tym algorytmów uczenia maszynowego, systemy te analizują dane kliniczne pacjentów, takie jak wyniki badań laboratoryjnych, obrazowania medycznego i historię chorób, aby wnioskować o prawdopodobieństwie występowania określonych stanów i schorzeń [20]. DDSS mogą być szczególnie użyteczne w skomplikowanych lub niejednoznacznych przypadkach, gdy wiele chorób może dawać podobne objawy lub gdy mamy do czynienia z rzadkimi chorobami, z którymi pracownicy ochrony zdrowia nie spotykają się na co dzień [21]. Istnieje kilka kluczowych elementów DDSS, w tym:

- Pozyskiwanie i wstępne przetwarzanie danych:
 - Polega na gromadzeniu danych o pacjencie z różnych źródeł, takich jak elektroniczna dokumentacja medyczna, obrazowanie medyczne, wyniki laboratoryjne i notatki kliniczne. Wstępne przetwarzanie danych polega na oczyszczeniu, normalizacji i przekształceniu surowych danych w odpowiedni format do dalszej analizy [22].
- Baza wiedzy:
 - Baza wiedzy zawiera wiedzę medyczną, taką jak objawy, czynniki ryzyka i kryteria diagnostyczne dla różnych chorób. Informacje te mogą pochodzić z wiedzy ekspertów, wytycznych klinicznych lub literatury badawczej [23].
- Silnik wnioskowania:
 - Silnik wnioskowania jest głównym elementem DDSS, odpowiedzialnym za przetwarzanie danych pacjenta i wiedzy medycznej w celu wygenerowania zaleceń diagnostycznych lub prognoz. Za logikę silnika odpowiadają zaawansowane algorytmy, w tym te oparte o metody uczenia maszynowego [24].

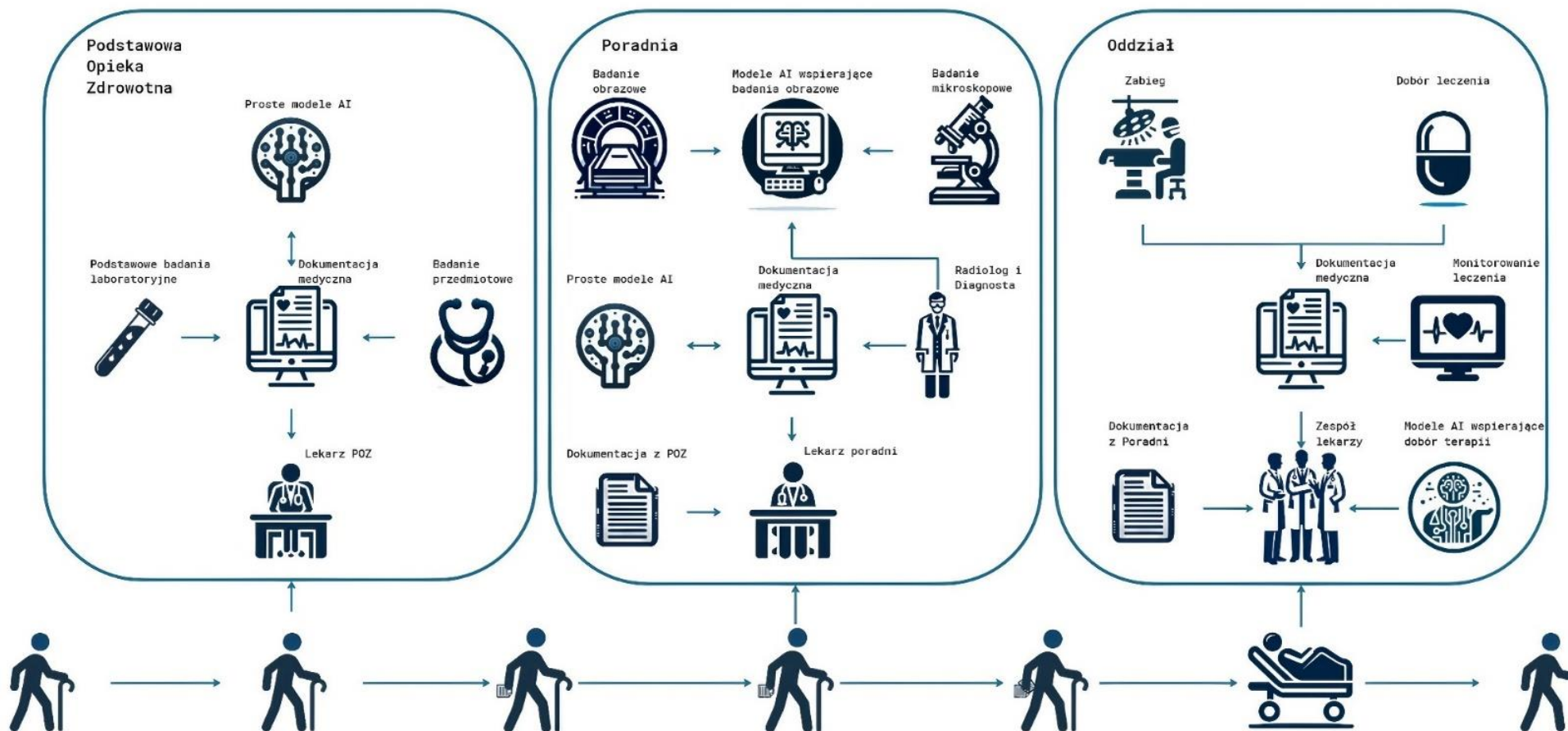
3.5. Systemy wsparcia decyzji diagnostycznych w ścieżce medycznej pacjenta

Rola systemów wsparcia decyzji diagnostycznych w kontekście ścieżki medycznej pacjenta rozpoczyna się od pierwszego kontaktu z systemem opieki zdrowotnej, przechodzi przez etapy diagnozy, planowania i realizacji leczenia, a kończy na rehabilitacji i edukacji zdrowotnej. DDSS mają potencjał do znaczącego ulepszenia każdego z tych etapów, zapewniając precyzyjniejszą diagnozę, personalizowane podejście terapeutyczne oraz efektywniejszą prewencję [5, 18, 25].

Rozpoczynając od etapu rejestracji i pierwszego kontaktu, DDSS mogą wspierać pracowników ochrony zdrowia w zebraniu i analizie podstawowych informacji o pacjencie, w tym papierowej historii medycznej zebranej z różnych placówek. W kolejnym kroku, podczas oceny stanu zdrowia i anamnezy, systemy te mogą pomóc w przedstawieniu dziesiątek stron historii medycznej w przystępnej formie, jej interpretacji, ocenie ogólnego stanu pacjenta oraz zaproponowaniu badań uzupełniających [26].

W fazie badań diagnostycznych, takich jak badania laboratoryjne czy obrazowe, DDSS mają kluczowe znaczenie w interpretacji wyników i wspieraniu lekarzy w wykrywaniu subtelnych anomalii, które mogą wskazywać na konkretne schorzenia [27].

Po postawieniu diagnozy, DDSS mogą wspierać lekarzy w opracowaniu planu leczenia, oferując dostosowane do indywidualnych potrzeb pacjenta rekomendacje terapeutyczne. W trakcie realizacji leczenia, te systemy pozwalają na monitorowanie reakcji pacjenta i dostosowywanie terapii, zapewniając tym samym lepsze wyniki leczenia [28]. Gdy leczenie pacjenta wymaga interwencji chirurgicznej, lekarze mają już do dyspozycji narzędzia wspomagane sztuczną inteligencją [29] (Rycina 3).



Rycina 3. Przykład ścieżki pacjenta z uwzględnieniem narzędzi AI w ochronie zdrowia

3.5.1. Podstawowa opieka zdrowotna

Podstawowa opieka zdrowotna stanowi fundament systemu zdrowia, zapewniając nie tylko bezpośrednią opiekę nad pacjentem, ale także będąc doskonałym źródłem danych o jego stanie zdrowia. Zwłaszcza w sytuacji, gdy pacjent korzysta przez wiele lat z usług jednej placówki bądź na bieżąco uzupełnia dokumentację medyczną, zgromadzone informacje mogą dostarczyć cennych wskazówek dotyczących jego stanu zdrowia.

W badaniach przeprowadzonych przez Bouha i wsp. podkreślono, jak AI może przyczynić się do digitalizacji danych medycznych [30]. Digitalizacja danych z poradni, klinik i innych placówek medycznych pozwala na ich szybsze przetwarzanie i lepszą dostępność. Badanie przeprowadzone przez zespół z Flatiron Health pokazuje, jak techniki NLP mogą być wykorzystane do efektywnego wydobywania informacji z nieustrukturyzowanych dokumentów medycznych, takich jak notatki kliniczne czy wyniki badań. Proces ten umożliwia przekształcanie danych w strukturalne zmienne, co ułatwia pracę zarówno lekarzom jak i pozwala na ich wykorzystanie w DDSS [31].

Interesującym badaniem mogącym mieć zastosowanie zwłaszcza w podstawowej opiece zdrowotnej jest diagnostyka pierwotnych niedoborów odporności (primary immunodeficiencies, PID). W badaniu opublikowanym w "The Journal of Allergy and Clinical Immunology" przedstawiono zastosowanie metody opartej na AI do oceny ryzyka PID na poziomie populacji. Wykorzystano do tego dwuetapowy system analityczny, który łączył analizator SPIRIT® firmy Jeffrey Modell Foundation z niestandardowym algorytmem uczenia maszynowego. Badanie przeprowadzono na niemal 430 000 osób, które były oceniane co 6 miesięcy przez 2.5 roku. Osoby o wysokim ryzyku były poddawane dalszej analizie przy użyciu modelu w celu ustalenia prawdopodobieństwa wystąpienia PID i priorytetyzacji skierowań do poradni immunologicznej [32].

Wreszcie, AI może odegrać kluczową rolę w proponowaniu badań laboratoryjnych. Zastosowanie algorytmów predykcyjnych do analizy danych medycznych może wskazać na potrzebę dodatkowych badań, które mogą być kluczowe dla wczesnej diagnozy. W badaniu Hoffmanna i wsp. przedstawiono zastosowanie technik uczenia maszynowego do generowania ścieżek diagnostycznych. W pracy pokazano zastosowanie prostych modeli drzew decyzyjnych, zbudowane na danych z badań pacjentów z przewlekłym wirusowym zapaleniem wątroby typu C. Wyniki pokazały, że modele te są skuteczne w tworzeniu prostych

i co ważniejsze - zrozumiałych ścieżek diagnostycznych, które mogą wspierać lekarzy w podejmowaniu decyzji dotyczących dalszych badań laboratoryjnych [33].

3.5.2. Środowisko szpitalne

W środowisku szpitalnym, sztuczna inteligencja przyczynia się do ulepszenia procesów takich jak triaż pacjentów, alertowanie o ciężkich stanach i chorobach wymagających szybkiego reagowania, ponadto ma zastosowanie w klasyfikacji chorób, w tym wyszukiwaniu podobnych przypadków medycznych w bazie szpitalnej.

Badania nad triażem pacjentów z użyciem AI w szpitalnych oddziałach ratunkowych (SOR) były intensywnie prowadzone w czasie pandemii COVID-19. Z badań przeprowadzonych przez Soltana i wsp. wynika, że AI może skutecznie przesiewać pacjentów z COVID-19, wykorzystując rutynowo zbierane dane takie jak wyniki badań krwi (morfologia, CRP, ALT, D-Dimer) i gazometrii, dostępnych zazwyczaj w ciągu pierwszej godziny od przybycia pacjenta do placówki [34]. W innym badaniu, opisanym przez Taylora i wsp. wykorzystano modele oparte o splotowe sieci neuronowe VGG, ResNet oraz Inception do retrospektywnej oceny RTG klatki piersiowej wykazując, że modele te mogą zapewnić wysokospecyficzne przesiewanie pacjentów z umiarkowanymi i ciężkimi stanami odmy opłucnowej [35].

Znaczący potencjał predykcji ciężkich stanów pacjentów z wykorzystaniem AI zbadano m.in. w wykrywaniu i leczeniu sepsy, szczególnie wśród noworodków i niemowląt w oddziałach intensywnej terapii noworodków. Zastosowanie modeli ML do wykrywania sepsy, opierających się na zmienności tętna i testach laboratoryjnych pozwoliło na wczesne ostrzeżenie o pogorszeniu stanu zdrowia pacjenta [36].

W innym badaniu, dotyczącym mikroangiopatii zakrzepowej (thrombotic microangiopathy, TMA) wykorzystano model TMA-INSIGHT (Thrombotic Microangiopathy Identification and Scoring for Optimal Guidance) oparty o algorytmy drzewiaste ze wzmocnieniem gradientowym XGBoost, LightGBM, do klasyfikacji etiologii TMA na podstawie danych klinicznych i wyników badań laboratoryjnych pacjentów. Algorytm ten wykazuje skuteczność w różnicowaniu pomiędzy różnymi rodzajami TMA, w tym cechującą się wysoką śmiertelnością zakrzepową płamicą małopłytkową (thrombotic thrombocytopenic purpura, TTP) a innymi typami TMA. Wzorce kliniczne i laboratoryjne, których model nauczył się do rozpoznawania pacjentów z TTP mogą pełnić kluczową rolę w alertowaniu podejrzenia choroby [37].

W ostatnich latach, w badaniach nad amyloidozą serca wykorzystywano metody uczenia maszynowego takie jak maszyny wektorów nośnych (support vector machine, SVM), algorytmy

drzewiaste [38] i splotowe sieci neuronowe [39]. Do różnicowania między amyloidozą a innymi chorobami kardiologicznymi używano danych z EKG, rezonansu magnetycznego oraz spektroskopii mas [40, 41].

AI było także szeroko stosowane w klasyfikacji nowotworów krwi. Głównym polem eksploracji naukowców była analiza zdjęć mikroskopowych leukocytów. Do tego celu wykorzystywano splotowe sieci neuronowe, zarówno modele oparte na powszechnie stosowanych architekturach VGG, ResNet, GoogLeNet [42–44], jak i własne architektury np. DeepLeukNet [45].

W badaniu przeprowadzonym przez Oh i wsp. zastosowano podejście do klasyfikacji chorób, wykorzystując sztuczną inteligencję do wyszukiwania podobnych przypadków medycznych w bazie szpitalnej. Ich badanie skupiało się na mierzeniu podobieństwa pacjentów z wieloma chorobami, wykorzystując zaawansowany model oparty na CNN. Model ten pozwalał na łączenie uczenia na podstawie danych pacjenta z uczeniem podobieństwa między danymi pacjentów. Wykorzystując dane z Narodowej Służby Ubezpieczeń Zdrowotnych Korei, zespół Oh zademonstrował model osiągający wysokie wyniki w porównaniu z istniejącymi modelami do mierzenia podobieństwa pacjentów z wieloma chorobami [46].

3.5.3. Laboratorium

W laboratorium diagnostycznym sztuczna inteligencja może być wbudowana w analizator diagnostyczny, stanowić osobne narzędzie dopełniające badanie oraz agregować wyniki w przystępny raport laboratoryjny.

W analizatorach diagnostycznych, takich jak CellaVision DC-1, sztuczna inteligencja odgrywa kluczową rolę w automatyzacji preklasyfikacji leukocytów. DC-1 wykorzystuje mikroskop, wysokiej jakości kamerę cyfrową oraz metody uczenia głębokiego do preklasyfikacji komórek z barwionych rozmazów krwi obwodowej. Dzięki temu, diagnosta może zweryfikować większą ilość komórek niż w standardowej metodzie mikroskopowej. Automatyczne rozwiązanie do klasyfikacji komórek krwi przyspiesza także czas realizacji badania oraz zmniejsza subiektywność i zmienność wyników między różnymi diagnostami [47]. Z kolei automatyzacja mikroskopii z UriSed 3 PRO znacznie usprawnia proces analizy moczu. To urządzenie wykorzystuje obrazy w kontraście fazowym i jasnym polu, uzyskane za pomocą kamery cyfrowej z 2,2 μ L odwirowanego moczu. Następnie modele oparte o sieci neuronowe klasyfikują obrazy krwinek, nabłonków oraz wałeczków [48].

W cytometrii przepływowej splotowe sieci neuronowe pozwoliły na osiągnięcie 99% skuteczności w klasyfikacji komórek takich jak THP1, MCF7, MB231, PBMC [49]. Ponad 95% skuteczność osiągnięto także w klasyfikacji nowotworów z komórek B wykorzystując algorytmy drzewiaste [50]. Clichet i wsp. poddali wyniki wieloparametrycznej cytometrii przepływowej selekcji cech metodą Boruta oraz użyli modelu regresji logistycznej ElasticNet do klasyfikacji pacjentów z zespołem mielodysplastycznym. Wysokie wyniki badania (pole pod krzywą – area under the curve, AUC = 0.935) pokazują, że przy dobrze przygotowanych danych proste algorytmy również mają zastosowanie [51].

W patomorfologii z wykorzystaniem AI można przeanalizować obraz całego slajdu. Naukowcy wykorzystywali splotowe sieci neuronowe do detekcji mitozy i jąder, klasyfikacji komórek i tkanek [52]. Interesującym badaniem była klasyfikacja zmian w biopsjach piersi w oparciu o klasy: zmiany nieproliferacyjne, proliferacyjne, nietypowy przerost przewodowy, rak przewodowy in situ oraz inwazyjny rak przewodowy. Badacze porównali skuteczność modelu (55%) do skuteczności oceny zmian przez 45 patomorfologów, którzy osiągnęli wynik 65%. W tym pokazano potencjał zastosowania CNN w patomorfologii jednak jak zauważyli

autorzy zaledwie 180 slajdów do nauki modelu może być niewystarczające w warunkach klinicznych [53].

3.5.4. Diagnostyka obrazowa

Diagnostyka obrazowa, znacznie ewoluowała dzięki zastosowaniu sztucznej inteligencji. Badania rentgenowskie (RTG), rezonans magnetyczny (MR), ultrasonografia (USG), mammografia oraz tomografia komputerowa (TK), były integrowane z metodami uczenia maszynowego wprowadzając większą precyzję i efektywność diagnostyki. Szczególnym okresem wzrostu zainteresowania AI w diagnostyce obrazowej był czas pandemii COVID-19, kiedy placówki medyczne dzieliły się danymi obrazowymi, a specjaliści od sztucznej inteligencji pracowali nad stworzeniem optymalnych modeli do klasyfikacji i triażu pacjentów z COVID-19 [54].

Ostatnie badania RTG skupiały się na klasyfikacji pacjentów z COVID-19 [54, 55]. AI znalazło również zastosowanie w detekcji patologii, takich jak kardiomegalia, niedodma, odma, wysięk, guzy, zapalenie płuc, obrzęk, rozedma płuc, zwłóknienie płuc i wątroby, przepuklina, oraz zwyrodnienie stawów [56].

Z kolei w rezonansie magnetycznym metody AI były eksplorowane głównie w obszarze detekcji nowotworów, w tym nowotworu piersi, prostaty i odbytnicy. Interesującym badaniem była przeprowadzona przez Shin i wsp. klasyfikacja 30 zmian przypominających glejaka przy użyciu metod uczenia głębokiego i obrazowania dyfuzyjnego, FLAIR (fluid-attenuated inversion recovery) oraz T1CE (T1-Post contrast-enhanced). Model zaproponowany przez autorów osiągnął podobne wyniki do radiologów, odpowiednio 72.3% i 72.6% przy 130 pacjentach w grupie testowej [57].

Wypracowane przez naukowców modele są coraz częściej transformowane w fizyczny produkt wspierający lekarzy w codziennej pracy. W latach 2016-2021 zarejestrowano w Agencji Żywności i Leków (Food and Drug Administration, FDA) produkty pomocne w ocenie ultrasonograficznej tarczycy, piersi, serca, tętnic szyjnych i pęcherza [58].

Warto zauważyć, że wdrożenie takich narzędzi nie ma na celu zastąpienia pracowników ochrony zdrowia, ale zapewnienie dodatkowej ochrony przed błędami poznawczymi wynikającymi m.in. z aktywizacji w umyśle heurystyki reprezentatywności[59]. Yoon i wsp. wykazali, że radiolodzy wspierani przez sztuczną inteligencję zwiększyli skuteczność wykrywania nowotworu piersi w badaniu mammograficznym z 80.1% do 88.1% [60].

Metody uczenia głębokiego mogą także ułatwiać diagnostykę przez polepszanie jakości obrazu np. w rekonstrukcji obrazów niskodawkowej tomografii komputerowej. Tradycyjne techniki rekonstrukcji, takie jak filtrowana projekcja wsteczna (filtered backpropagation, FBP) i rekonstrukcja iteracyjna (iterative reconstruction, IR), są ograniczone przez długie czasy obliczeń i niską jakość obrazu, szczególnie w przypadku niskodawkowej TK. Metody uczenia głębokiego, takie jak ResNet, U-Net i generatywne sieci antagonistyczne (generative adversarial network, GAN), znacząco podniosły jakość obrazów, redukując szумы i artefakty przy jednoczesnym zachowaniu struktur anatomicznych [61]. FDA zaaprobowała systemy rekonstrukcji obrazów oparte na głębokim uczeniu, takie jak Advanced intelligent ClearIQ Engine (AiCE) [62] i GE Healthcare's TrueFidelity [63], które wykazują lepsze wyniki niż konwencjonalne metody

Pomimo zalet modeli uczenia głębokiego, ich wytłumaczalność stanowi spore wyzwanie, szczególnie w kontekście decyzji diagnostycznych. Zwłaszcza w ochronie zdrowia ważne jest, aby narzędzie wskazywało nie tylko wynik, ale również przedstawiło przejrzysty proces podejmowania decyzji. Głębokie architektury sieci neuronowych, choć są potężnym narzędziem analitycznym, często brak im przejrzystości w wyjaśnianiu podstaw decyzji modelu. Rozwijająca się dziedzina eksplikowalnej sztucznej inteligencji, skupiona wokół zwiększania wiarygodności i interpretowalności tych metod, może w przyszłości umożliwić i przyspieszyć wdrażanie sztucznej inteligencji w praktyce klinicznej [64].

3.5.5. Planowanie i monitorowanie leczenia

Zdolność sztucznej inteligencji do analizy dużych zbiorów danych, identyfikacji wzorców i predykcji otworzyło nowe horyzonty w medycynie precyzyjnej - indywidualnym planowaniu i monitorowaniu leczenia. Zastosowanie AI w medycynie nie ogranicza się jedynie do poprawy efektywności diagnostycznej, ale również obejmuje personalizację terapii, dostosowując leczenie do unikalnych cech pacjenta. Dzięki rozwojowi uczenia głębokiego, możliwe jest identyfikowanie najskuteczniejszych metod leczenia dla każdego pacjenta. W ostatnich latach ten obszar badawczy był szczególnie eksplorowany w onkologii.

AI znalazło zastosowanie w genomice, co umożliwiło dostosowanie leczenia do konkretnego pacjenta. Przykładem jest leczenie medulloblastomy u dzieci, gdzie dzięki zastosowaniu AI do analizy setek eksomów uzyskano informacje o odpowiednim leczeniu

i dawce dla pacjentów, co znacznie zmniejszyło ryzyko powikłań neurokognitywnych i wtórnych nowotworów [65]. AI zostało także wykorzystane do analizy danych genetycznych pacjentów, w celu przewidywania reakcji na radioterapie i chemioterapie. W badaniu Dubey i wsp. dotyczącym zastosowania sztucznej inteligencji w planowaniu leczenia raka piersi, naukowcy podzielili proces na dwa etapy mające na celu optymalizację i personalizację terapii. W pierwszym etapie zastosowano proste algorytmy uczenia maszynowego, takie jak regresja logistyczna i lasy losowe, do oceny stanu pacjenta po 5 latach od rozpoczęcia leczenia. Celem tego etapu było wykorzystanie klasyfikatora, który jest zarówno dokładny, jak i wyjaśnialny. Drugi etap skupiał się na badaniu alternatywnych planów leczenia za pomocą algorytmu k najbliższych sąsiadów (K-Nearest Neighbors, KNN). W tym etapie, statystyki dotyczące "sąsiednich" pacjentów zostały wykorzystane do zbadania możliwości leczenia, takich jak konieczność chemioterapii, optymalna sekwencja radioterapii w połączeniu z zabiegiem chirurgicznym. Celem drugiego etapu było wsparcie lekarza w wyborze terapii, która maksymalizuje szanse na przeżycie po 5 latach [28].

Do innowacji w zakresie integracji sztucznej inteligencji w praktyce chirurgicznej należą operacje z asystą robotów oraz systemy nawigacji oparte na rzeczywistości rozszerzonej (augmented reality, AR) Chirurgia wspomagana robotami prowadzi do mniej inwazyjnych zabiegów, zwiększa precyzję, skraca czas rekonwalescencji i minimalizuje powikłania. Technologie rzeczywistości rozszerzonej, dostarczając wsparcie nawigacyjne i wizualizacje w czasie rzeczywistym, umożliwiają chirurgom, zwłaszcza tym mniej doświadczonym, wykonywanie skomplikowanych zabiegów z większą dokładnością i bezpieczeństwem, demokratyzując tym samym dostęp do wysokiej jakości opieki chirurgicznej [29].

3.6. Duże modele językowe w systemach wsparcia decyzji klinicznych

Od 2022 roku świat sztucznej inteligencji skupiał się głównie na generatywnej sztucznej inteligencji. Zaliczamy do niej modele generujące tekst, obrazy i dźwięk. Modele te znacząco różnią się od opisywanych wcześniej wielkością zbiorów treningowych. Czas trenowania dużego modelu językowego może w zależności do liczby parametrów wynosić od kilku dni do kilku miesięcy. Koszt trenowania modelu z 7 miliardami parametrów np. Llama 2 7B to koszt około 85 tysięcy dolarów [66], a modelu z 175 miliardami parametrów np. GPT 3 to koszt około 5 milionów dolarów [67]. Modele te pozwalają na stworzenie automatycznych systemów redagowania artykułów, narzędzi wspomagających pisanie scenariuszy czy chatbotów o zaawansowanych umiejętnościach komunikacyjnych. LLM mogą także pomagać w tworzeniu treści edukacyjnych, spersonalizowanych wiadomości marketingowych, a nawet odpowiedzi w systemach wsparcia użytkowników [68].

Duże modele językowe w ochronie zdrowia stanowią znaczący postęp w porównaniu z wcześniejszymi technologiami NLP, oferując lepsze zrozumienie kontekstu i złożonych relacji występujących w języku medycznym. W przeciwieństwie do tradycyjnych systemów NLP, które skupiały się na podstawowym przetwarzaniu tekstu i często wymagały ręcznego definiowania reguł, duże modele językowe uczą się na ogromnych zbiorach danych, co pozwala na wypracowanie bardziej zaawansowanego zrozumienia języka [69]. Na przykład, model GatorTron, uczony na ponad 90 miliardach słów z dokumentacji medycznej, wykazał duży potencjał do analizowania historii medycznych pacjentów i uzupełniania wypisów pacjentów na podstawie dokumentacji z hospitalizacji [70].

LLM mogą być wykorzystywane do analizy historii medycznych pacjentów, a co istotniejsze umożliwić analizę danych z nieustrukturyzowanych źródeł, w tym notatek lekarskich i historii medycznej bez potrzeby zaawansowanych narzędzi strukturyzujących te dane. Ponadto, duże modele językowe mają potencjał do ulepszania komunikacji między lekarzami a pacjentami, przetwarzając informacje medyczne na bardziej zrozumiały dla pacjentów język, co jest kluczowe w zwiększaniu ich zaangażowania w proces leczenia i poprawie zrozumienia stanu zdrowia oraz dostępnych opcji terapeutycznych [69].

3.7. Wytłumaczalność modeli sztucznej inteligencji

Wytłumaczalność sztucznej inteligencji (explainable artificial intelligence, xAI) w medycynie jest kluczowa dla zapewnienia transparentności procesów decyzyjnych. Decyzje dotyczące diagnostyki i leczenia, podjęte przez modele AI, muszą być wytłumaczalne ze względu na wysokie ryzyko poważnych konsekwencji w przypadku błędów. xAI oznacza odpowiedź na pytanie – jakie cechy wzięte przez model sztucznej inteligencji do analizy były najistotniejsze do klasyfikacji, bądź predykcji. Jasne zrozumienie mechanizmów, na podstawie których AI podejmuje decyzje, jest fundamentem dla budowania zaufania między pracownikami ochrony zdrowia, a technologią [71].

Prostota architektury modelu AI często koreluje z większą wytłumaczalnością. Modele takie jak regresja logistyczna czy drzewa decyzyjne są bardziej przejrzyste w porównaniu do skomplikowanych głębokich sieci neuronowych. Mimo to, istnieją zaawansowane metody, takie jak Local Interpretable Model-agnostic Explanations (LIME) [72] i SHapley Additive exPlanations (SHAP) [73], które umożliwiają wyjaśnienie nawet skomplikowanych decyzji AI, identyfikując kluczowe cechy wpływające na ich wyniki. xAI umożliwia lekarzom głębsze zrozumienie rekomendacji wygenerowanych przez AI, przyczyniając się także do eksploracji nowych sposobów postrzegania problemu rozwiązywanego przez AI. Wytłumaczalność sztucznej inteligencji w medycynie przekracza ramy technologiczne, stając się istotnym elementem etycznym i prawnym, wspierającym transparentność i odpowiedzialność w opiece zdrowotnej [74].

3.8. Systemy wsparcia decyzji diagnostycznych w chorobach rzadkich

Systemy wsparcia decyzji diagnostycznych są szczególnie istotne w chorobach rzadkich, gdzie specyfika i rzadkość objawów stanowią wyzwania diagnostyczne. Zgodnie z danymi Narodowego Instytutu Zdrowia (national institute of health, NIH), zidentyfikowano ponad 7000 chorób rzadkich, co sprawia, że ich rozpoznanie może być wyzwaniem nawet dla doświadczonych specjalistów [75]. Wiele chorób rzadkich manifestuje się nieswoistymi objawami i zanim wzbudzą obawy lekarza, mogą poczynić znaczne szkody w organizmie pacjenta. Proces diagnozowania choroby rzadkiej jest przez to długi i wyczerpujący. U 25% pacjentów, diagnoza jest stawiana między 5 a 30 lat po pojawieniu się pierwszych objawów i wymaga udziału kompleksowego zespołu klinicystów [76]. Modele uczenia maszynowego mogą skrócić ścieżkę diagnostyki pacjentów z chorobami rzadkimi. Wilson i wsp. użyli algorytmu Light Gradient Boosting Machine (LightGBM) do wytrenowania modelu do diagnostyki choroby Gauchera (Gaucher disease, GD) i porównali go z praktyką kliniczną. Wytrenowany model był około 10-20 razy skuteczniejszy w identyfikacji pacjentów z GD niż kliniczny algorytm diagnostyczny [21].

Kluczowe dla rozwoju systemów DDSS dla chorób rzadkich jest współpraca między lekarzami specjalistami a inżynierami uczenia maszynowego. Z uwagi na ograniczoną ilość dostępnych danych treningowych, istotne jest, aby wiedza medyczna specjalistów była integralną częścią narzędzi AI. W tym kontekście, kluczowe staje się budowanie współpracy i dzielenie się danymi i wiedzą między różnymi ośrodkami. Między innymi w odpowiedzi na tę potrzebę, zainicjowano tworzenie baz wiedzy, które gromadzą specjalistyczne dane na temat chorób rzadkich. Dzięki inicjatywom takim jak Orphanet [77], które wypracowało sobie pozycję lidera w dostarczaniu informacji o tych chorobach, zwiększono możliwość prowadzenia badań nad rzadkimi chorobami. Orphanet jest cennym zasobem danych wykorzystywanych przez systemy DDSS. Przykłady systemów zaprojektowanych specjalnie dla rzadkich zaburzeń, takie jak FindZebra [78] i Ada DX [79], demonstrują, jak wiedza zdeponowana w bazach danych może wspierać proces diagnostyczny.

3.8.1. Choroba Andersona-Fabry'ego

Choroba Andersona-Fabry'ego (Anderson-Fabry disease, AFD) jest ultra rzadką, sprzężoną z chromosomem X wynikającą z niedoboru enzymu alfa-galaktozydazy A. Niedobór ten prowadzi do gromadzenia się glikosfingolipidów, zwłaszcza globotriaosyloceramidu (Gb3), w różnych tkankach i narządach. AFD objawia się szerokim spektrum symptomów, w tym bólem kończyn, naczynekami, hipohydrozą, problemami żołądkowo-jelitowymi, utratą słuchu oraz postępującymi powikłaniami nerkowymi, sercowymi i mózgowo-naczyniowymi. Częstość występowania AFD szacuje się na 1 na 40 000 do 1 na 170 000 urodzeń, we wszystkich grupach etnicznych. Ze względu na postać heterogenną i rzadkość występowania choroby, jest ona często niedostatecznie lub błędnie diagnozowana, co prowadzi do opóźnień w leczeniu i większego ryzyka powikłań [80, 81].

W badaniu Jefferies i wsp. wykorzystali zbiór danych dokumentacji medycznej 4 978 pacjentów z AFD oraz 1 000 000 pacjentów bez informacji o chorobie w dokumentacji medycznej do wytrenowania modelu sztucznej inteligencji odróżniającego pacjentów z AFD od pacjentów z innymi chorobami. Naukowcy nie wskazują algorytmu jaki użyli do wytrenowania modelu, niemniej osiągnęli wysoką skuteczność klasyfikacji pacjentów w grupie testowej AUC = 0.82 [3].

4. Cel pracy

Wzrost popularności sztucznej inteligencji pomaga naukowcom i lekarzom wdrażać narzędzia oparte o metody uczenia maszynowego w praktykę kliniczną. Pozytywne wyniki tych wdrożeń zwiększają zaufanie pracowników ochrony zdrowia do sztucznej inteligencji, jednak za samą technologią muszą stać solidne fundamenty etyczno-prawne. Medycy powinni nauczyć się korzystać z tych narzędzi, a ich twórcy zadbać o to, żeby były transparentne i interpretowalne.

Celem pracy jest:

- wdrożenie prototypu systemu wsparcia decyzji diagnostycznych dla choroby Andersona-Fabry'ego wraz z opisaniem metodologii tworzenia narzędzia wykorzystującego przetwarzanie języka naturalnego jako sposób ekstrakcji danych klinicznych z dokumentacji medycznej oraz algorytmu punktowego różnicującego pacjentów z chorobą Andersona-Fabry'ego, opracowanego przez specjalistów zajmujących się diagnostyką tej choroby,
- zastosowanie algorytmu Shapley Additive Explanations jako metody wytłumaczalności modeli splotowych sieci neuronowych do klasyfikacji pojedynczych zdjęć mikroskopowych leukocytów.

Poprzez realizację tych badań, niniejsza praca ma wnieść wkład do rosnącej liczby dowodów dotyczących zastosowania uczenia maszynowego w systemach wsparcia decyzji diagnostycznych w praktyce i zapewnienie kompleksowego zrozumienia potencjalnych korzyści, wyzwań i ograniczeń związanych z ich stosowaniem. Wyniki tego badania mogą przyczynić się do opracowania i wdrożenia bardziej skutecznych, wytłumaczalnych strategii diagnostycznych wykorzystujących sztuczną inteligencję, ze szczególnym uwzględnieniem diagnostyki chorób rzadkich, w tym choroby Andersona-Fabry'ego.

5. Omówienie prac włączonych do cyklu

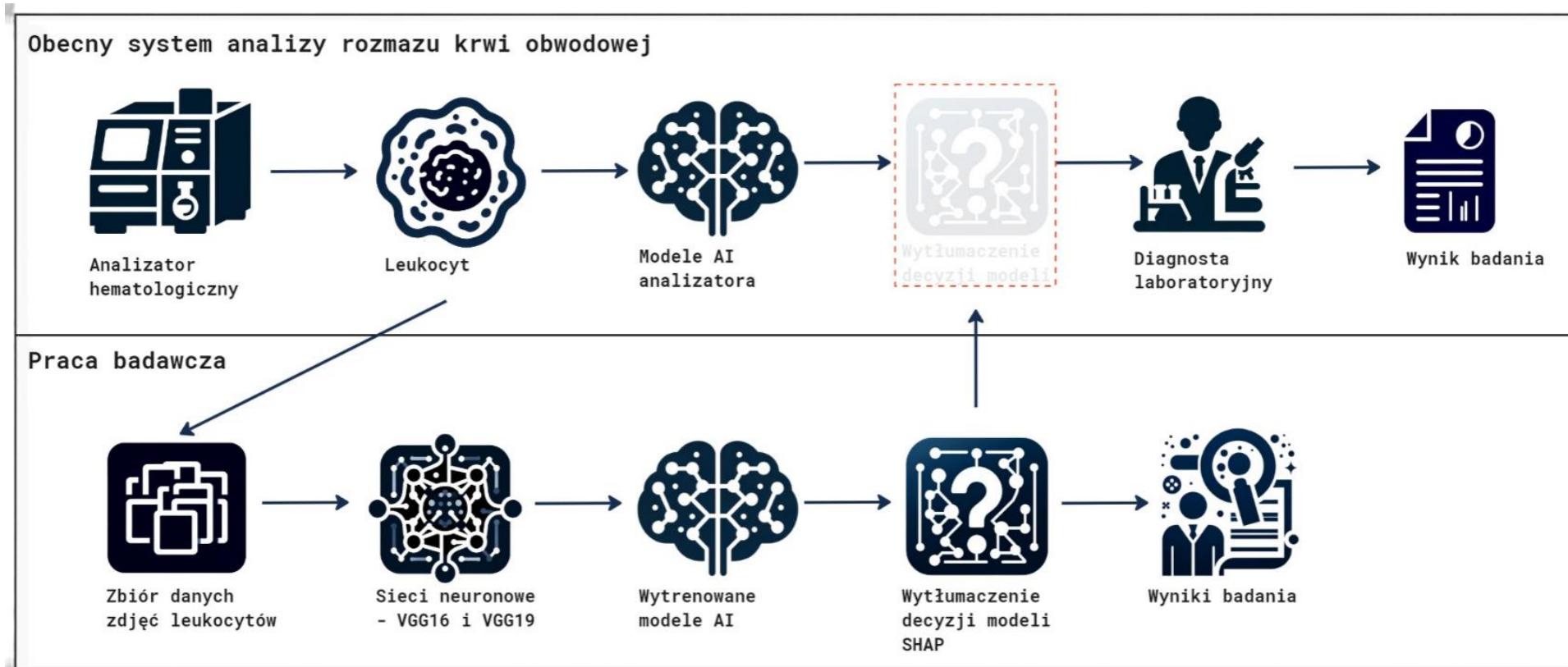
5.1. Artykuł oryginalny I

Michalski A, Duraj K, Kupcewicz, B. *Leukocyte deep learning classification assessment using Shapley additive explanations algorithm*. Int J Lab Hematol. 2023; 1-6. <https://doi.org/10.1111/ijlh.14031> (Załącznik I)

Mikroskopowa ocena rozmazu krwi obwodowej, jest jednym z podstawowych badań w diagnostyce laboratoryjnej. Badanie to polega na ilościowej i jakościowej analizie elementów morfotycznych. Jednym z jego elementów jest różnicowanie białych krwinek. Tradycyjnie metody klasyfikacji leukocytów opierają się na ich zliczaniu w obrazie mikroskopowym z i zliczaniu na sumatorach hematologicznych, co jest procesem czasochłonnym i podatnym na błędy. Pojawienie się splotowych sieci neuronowych otworzyło nowe możliwości automatyzacji tego procesu, oferując zwiększoną wydajność różnicowania krwinek przy zachowaniu wysokiej skuteczności. Jednak trudności w wytłumaczeniu decyzji modeli uczenia głębokiego stanowią poważne wyzwanie. Nietransparentny proces klasyfikacji komórek mimo wysokiej skuteczności utrudnia diagnostom laboratoryjnym zaufanie takim narzędziom.

Aby wypełnić tę lukę, obiecującym rozwiązaniem okazało się zastosowanie narzędzi wyjaśniających, takich jak SHapley Additive exPlanations (SHAP) DeepExplainer. Ten algorytm zapewnia wgląd w procesy decyzyjne modeli uczenia głębokiego poprzez identyfikację cech, które mają największy wpływ na przewidywania modelu. Nie tylko zwiększa to wiarygodność i przejrzystość zautomatyzowanych systemów klasyfikacji leukocytów, ale także pomaga w dalszym zrozumieniu istotnych cech morfologicznych różnicujących białe krwinki.

Niniejsze badanie ma na celu przedstawienie zastosowanie wstępnie wytrenowanych modeli splotowych sieci neuronowych - VGG16 i VGG19, wzbogaconych o SHAP DeepExplainer, do różnicowania leukocytów z obrazów rozmazów krwi obwodowej. Wykorzystując możliwości tych modeli i narzędzia wyjaśniającego, w badaniu przedstawiono wiarygodny system służący do klasyfikacji leukocytów.



Rycina 4. Schemat integracji metody wyjaśnialności modeli sztucznej inteligencji w automatycznym systemie analizy rozmazów krwi obwodowej.

wskazania istotnych elementów komórek do ich odróżnienia. Taki system mógłby pomóc diagnostom laboratoryjnym w weryfikacji dokładności klasyfikacji komórek i uzyskaniu głębszego wglądu w zasady działania automatycznych analizatorów hematologicznych (Rycina 4).

Podstawą badania jest zbiór danych obejmujący 17 092 obrazów rozmazów krwi obwodowej, publicznie udostępniony przez Acevedo i wsp. Zbiór danych obejmuje kilka typów leukocytów. W badaniu skoncentrowano się na pięciu podstawowych: neutrofilach, limfocytach, monocytach, eozynofilach i bazofilach. Do klasyfikacji komórek użyto wytrenowane splotowe sieci neuronowe - VGG16 i VGG19, poprzez uczenie transferowe, pozwalające na dostosowanie wstępnie wytrenowanych modeli do specjalistycznego zadania klasyfikacji leukocytów.

Wybór modeli VGG16 i VGG19 wynikał z ich udowodnionej skuteczności w zadaniach rozpoznawania obrazów. Stosując strategię uczenia transferowego, modele zostały dostosowane do konkretnego kontekstu klasyfikacji leukocytów. Kluczowym aspektem metodologii jest włączenie SHAP DeepExplainer, narzędzia zaprojektowanego w celu umożliwienia wytłumaczenia decyzji modeli uczenia głębokiego. SHAP DeepExplainer został użyty do analizy procesów decyzyjnych modeli, identyfikując kluczowe elementy komórkowe, które wpływają na wyniki klasyfikacji. Zbadano dwie różne strategie uczenia: zamrożenie warstw ekstrakcji cech CNN w celu udoskonalenia tylko końcowych warstw sieci neuronowej oraz uczenie całej sieci, aby zweryfikować czy sposób trenowania modelu ma wpływ na cechy różnicujące komórki. Modele zostały zoptymalizowane przy użyciu optymalizatora Adam, z malejącą szybkością uczenia.

Modele oparte o VGG16 i VGG19 osiągnęły wysoką dokładność przekraczającą 99% w obu strategiach treningowych. Ten wysoki poziom dokładności podkreśla potencjał modeli uczenia głębokiego do bezpiecznego zastosowania w praktyce laboratoryjnej, niemniej użyty zbiór danych zawierał dobrze przygotowane zdjęcia pojedynczych leukocytów bez komórek patologicznych i artefaktów występujących w rzeczywistych preparatach co miało wpływ na dokładność modeli. SHAP DeepExplainer umożliwił wgląd w proces decyzyjny modeli. W przypadku wstępnie wytrenowanych modeli, analiza SHAP podkreśliła znaczenie konturów komórek i jąder w różnicowaniu komórek. Natomiast, w przypadku w pełni wytrenowanych modeli nacisk przesunął się w kierunku obszaru cytoplazmy. Odkrycia te są ważne z dwóch

powodów. Po pierwsze, potwierdzają skuteczność SHAP DeepExplainer w odkrywaniu istotnych cech dla decyzji modeli, zwiększając w ten sposób wyjaśnialność i wiarygodność narzędzi AI wykorzystujących splotowe sieci neuronowe w diagnostyce mikroskopowej. Po drugie, zróżnicowany nacisk na elementy komórek między częściowo i w pełni wytrenowanymi modelami zapewnia cenną informację o wpływie strategii trenowania modeli na wybór istotnych cech do różnicowania komórek. Wykorzystanie metod umożliwiających wyjaśnianie decyzji modeli uczenia głębokiego ma znaczące implikacje dla diagnostyki laboratoryjnej. Dzięki zrozumieniu podstaw klasyfikacji modeli, diagności laboratoryjni mogą z większą pewnością interpretować i korygować wyniki automatycznej klasyfikacji.

Badanie to pokazuje potencjał połączenia splotowych sieci neuronowych z SHAP DeepExplainer do różnicowania białych krwinek. Ułatwiając konwergencję zaawansowanych metodologii uczenia maszynowego i wymagań aplikacji diagnostycznych, niniejsze badanie wzmacnia inicjatywy mające na celu wdrażanie narzędzi sztucznej inteligencji w ochronie zdrowia, zwiększając precyzję, skuteczność i zrozumiałość procedur diagnostycznych.

5.2. Artykuł oryginalny II

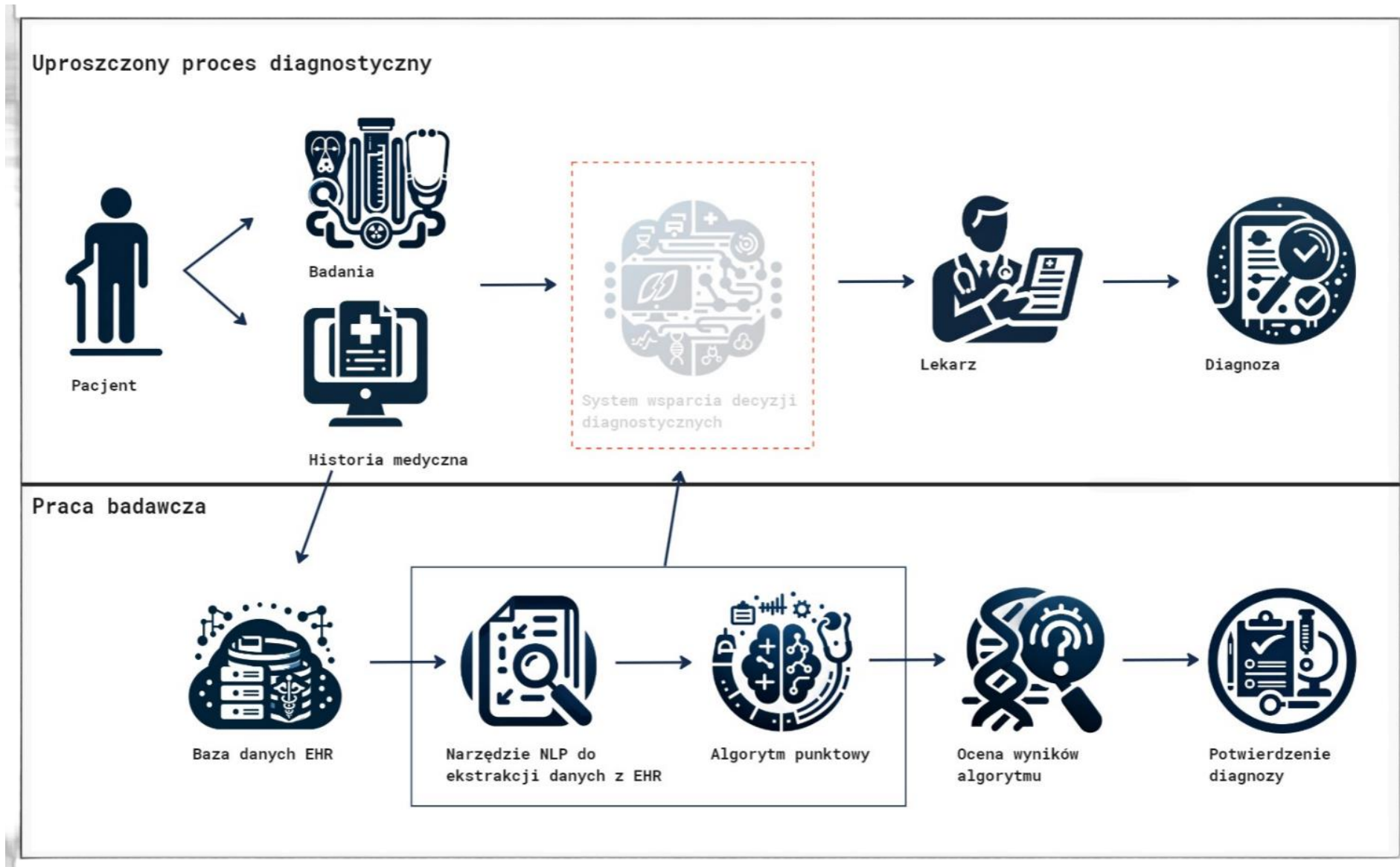
Michalski A, Lis K, Stankiewicz J, Kloska M.S, Dudziński M, Muras-Szwedziak K, Nowicki M, Bazan-Socha S, Dąbrowski J.M, Basak W.G. ***Supporting the diagnosis of Fabry disease using a natural language processing-based approach***. J. Clin. Med. 2023, 12(10), 3599; <https://doi.org/10.3390/jcm12103599> (Załącznik II)

Choroba Andersona-Fabry'ego jest ultrarzadką chorobą genetyczną. Zaburzenie to wynika z patologicznych wariantów genu GLA odpowiedzialnego za kodowanie enzymu lizosomalnego alfa-galaktozydazy A. Niedobór tego enzymu prowadzi do gromadzenia się glikosfingolipidów, zwłaszcza globotriaosylceramidów w lizosomach, powodując degenerację komórek, zapalenie tkanek i w konsekwencji dysfunkcję narządów. Jako zaburzenie sprzężone z chromosomem X, AFD zwykle objawia się poważniej u mężczyzn niż u kobiet, z objawami często pojawiającymi się we wczesnym dzieciństwie, w tym bólem neuropatycznym, zaburzeniami żołądkowo jelitowymi, upośledzoną potliwością i nietolerancją skrajnych temperatur.

Pomimo szerokiego spektrum objawów, mają one niespecyficzny charakter, dlatego rozpoznanie AFD jest często opóźnione o 7-10 lat od wystąpienia pierwszych objawów. To opóźnienie stwarza potrzebę wdrożenia metod ułatwiających wczesną diagnostykę w celu zapobieżenia nieodwracalnym uszkodzeniom narządów poprzez wdrożenie odpowiedniego leczenia. Integracja sztucznej inteligencji w procesach przesiewowych dokumentacji medycznej pacjentów jest obiecującym podejściem do przyspieszenia diagnostyki AFD.

Niniejsze badanie ma na celu opracowanie i walidację systemu wspierającego diagnostykę choroby Andersona-Fabry'ego opartego na punktacji cech klinicznych wskazanych przez lekarzy specjalistów. System ten polega na automatycznym wydobywaniu informacji z EHR przy użyciu metod przetwarzania języka naturalnego oraz obliczaniu ryzyka występowania AFD na podstawie zebranych danych o pacjencie (Rycina 5.)

W badaniu wykorzystano zanonimizowane dane medyczne pochodzące z bazy danych Saventic Health. Dane były przechowywane w bazie danych PostgreSQL, obejmowały szeroki zakres ustrukturyzowanych i nieustrukturyzowanych informacji,



Rycina 5. Schemat integracji sztucznej inteligencji w proces diagnostyczny

takich jak dane demograficzne (wiek i płeć), historia medyczna, objawy, wyniki badań laboratoryjnych, diagnozy, kody ICD-10, opisy obrazowania i epikryzy. Ekstrakcję i wstępne przetwarzanie danych przeprowadzono przy użyciu biblioteki SpaCy w języku programowania Python.

Po przeprowadzeniu analizy dokumentacji medycznej pacjentów przez stworzony przesiewowy system AFD, koordynatorzy medyczni w odpowiednich szpitalach zostali poinformowani o wynikach, w szczególności o pacjentach z wysokim ryzykiem choroby. Koordynatorzy ci, w porozumieniu z lekarzami prowadzącymi, określili konieczność wykonania testu aktywności alfa-galaktozydazy A.

Grupę badawczą stanowiło 13 pacjentów, u których wcześniej zdiagnozowano AFD (potwierdzone testem DBS). Dane tych pacjentów pochodziły z siedmiu różnych szpitali. Aby zasymulować reprezentatywną demografię szpitalną, utworzono grupę kontrolną, włączając 19 372 pacjentów z jednego szpitala z różnych oddziałów i poradni. Grupa kontrolna została sprawdzona pod kątem występowania w EHR wzmianek o AFD. Żaden z pacjentów nie miał w dokumentacji medycznej informacji o podejrzeniu tej choroby, natomiast ważne jest, aby zwrócić uwagę na możliwość wystąpienia niezdiagnozowanych przypadków AFD w tej grupie. Pacjenci w obu grupach byli odfiltrowani przez kryteria włączenia i wykluczenia. Wykorzystano dane pacjentów w wieku 18-75 lat. Kryteria wykluczenia eliminowały pacjentów pod opieką paliatywną, uzależnionych od alkoholu lub ze zdiagnozowanymi rozszianymi nowotworami lub poddawanych chemioterapii.

Do przesiewu pacjentów z podejrzeniem AFD użyto algorytmu punktowego. Trzydzieści punktowanych cech klinicznych zostało wybranych w oparciu o wiedzę lekarzy specjalistów doświadczonych w diagnostyce tej choroby. Punkty dla poszczególnych cech zostały przypisane w skali od zera do trzech, odzwierciedlając znaczenie diagnostyczne cechy, jej swoistość oraz wiek pacjenta. Cecha powszechna w populacji ogólnej w określonym wieku otrzymała niższą punktację, podczas gdy cechy nietypowo występujące u młodszych pacjentów - takie jak udar, zawał mięśnia sercowego, niewydolność nerek, choroby oczu i utrata słuchu - otrzymały wyższą punktację, wskazując na większe znaczenie dla ryzyka AFD. Informacje o cechach klinicznych pochodziły z kodów ICD-10 oraz z nieustrukturyzowanej dokumentacji medycznej, wydobyte przy użyciu metod przetwarzania języka naturalnego.

Skuteczność tych metod zależała od dokładnego wykrywania terminów medycznych i wymagała analizy setek opisów w celu udoskonalenia precyzji algorytmu.

Na podstawie wydobytych cech klinicznych, wszystkim pacjentom przydzielono punkty. Zbadano kilka progów odcięcia, w celu rozróżnienia pacjentów z AFD. Analiza grupy kontrolnej wykazała, że żaden z pacjentów nie otrzymał poniżej 3 punktów, dlatego ta wartość została wybrana jako próg odcięcia. Pacjenci zostali podzieleni na grupy niskiego i wysokiego ryzyka na podstawie przydzielonych punktów. Dokumentacja pacjentów wysokiego ryzyka została przeanalizowana przez dwóch lekarzy, którzy zdecydowali o konieczności powiadomienia koordynatora szpitalnego o podejrzeniu AFD.

Łącznie zidentyfikowano 92 pacjentów z czynnikiem ryzyka ≥ 4 . Wśród nich 80 należało do grupy kontrolnej. Lekarze przeglądający dokumentację medyczną tych pacjentów zasugerowali dalszą diagnostykę 15, w wyniku czego zdiagnozowano AFD u jednej osoby początkowo sklasyfikowanej w grupie kontrolnej z wysokim czynnikiem ryzyka. Pacjent ten miał 45 lat, w wywiadzie niewyjaśnioną kardiomiopatię, zawroty głowy, bezobjawowe zmiany niedokrwienne w MRI mózgu i przeszczep nerki u krewnego pierwszego stopnia.

Tworzenie systemów wsparcia decyzji diagnostycznych dla chorób rzadkich związane jest z wyzwaniem w postaci niewielkich rozmiarów zbiorów danych i wymagają alternatywnych strategii. W niniejszym badaniu przedstawiono podejście oparte na cechach klinicznych zdefiniowanych przez lekarzy specjalistów, zapewniając transparentność modelu i łatwość interpretacji. Integracja metod uczenia maszynowego z wiedzą ekspercką w system wsparcia decyzji diagnostycznych, oferuje znaczący postęp w identyfikacji chorób takich jak AFD. Objawy takie jak angiokeratoma, ból kończyn, nawracające gorączki i zaburzenia potliwości mogą być niezauważane lub ich znaczenie bagatelizowane przez wiele lat, co podkreśla znaczenie szczegółowe analizowanie dokumentacji medycznej. W związku z tym włączenie modeli sztucznej inteligencji do praktyki klinicznej poprzez analizę przesiewową dokumentacji pacjentów może usprawnić proces diagnostyczny, kierując osoby z grupy wysokiego ryzyka do dalszych badań. Ponadto modele działające w tle systemu szpitalnego i raportujące pacjentów z wysokim ryzykiem chorób rzadkich mogą zwiększyć świadomość tych chorób wśród pracowników, stanowiąc element edukacyjny systemu.

Badanie to pokazuje, że nieskomplikowane narzędzia oparte o metody przetwarzania języka naturalnego oraz ekspercka baza wiedzy mogą posłużyć do stworzenia skutecznego systemu wsparcia decyzji diagnostycznych w chorobie Andersona-Fabry'ego.

6. Wnioski

Digitalizacja systemu ochrony zdrowia, postęp medycyny oraz nowe metody sztucznej inteligencji pozwalają na stworzenie narzędzi, które mogą zmniejszyć obciążenie ochrony zdrowia przez automatyzację rutynowych zadań, przyspieszyć rozpoznanie chorób trudnych do zdiagnozowania, wspomóc personel medyczny w dostosowaniu terapii oraz zwiększyć precyzję zabiegów chirurgicznych.

W niniejszej rozprawie doktorskiej zbadano potencjał systemów wsparcia decyzji diagnostycznych działających w oparciu o metody sztucznej inteligencji. Skupiono się na dwóch problemach – wyjaśnialności modeli splotowych sieci neuronowych w klasyfikacji zdjęć mikroskopowych leukocytów z rozmazów krwi obwodowej oraz systemie przesiewowym dla choroby Fabry’ego wykorzystującym metody przetwarzania języka naturalnego do ekstrakcji cech klinicznych z nieustrukturyzowanych danych medycznych. Wyniki przeprowadzonych badań pozwalają na wyciągnięcia następujących wniosków:

- badanie wytłumaczalności metodą SHAP wytrenowanych splotowych sieci neuronowych VGG16 i VGG19 do klasyfikacji leukocytów wykazało, że sposób ich różnicowania jest związany z obszarem cytoplazmy, obrysem komórki i jądra komórkowego,
- modele uczenia głębokiego stworzone do klasyfikacji zdjęć mikroskopowych leukocytów charakteryzują się wysoką skutecznością, jednak wdrażanie ich w systemach diagnostycznych wymaga przemyślanej integracji metod pozwalających na wyjaśnienie decyzji modeli,
- zdiagnozowanie choroby Andersona-Fabryego u pacjenta wskazanego przez algorytm przesiewowy pokazuje potencjał stworzonego systemu. Wiedza naukowców i lekarzy specjalistów przepisana na algorytmy punktowe, w połączeniu z metodami przetwarzania języka naturalnego może stanowić skuteczne i transparentne narzędzie przesiewowe w chorobie Andersona-Fabry’ego,
- choroba Andersona-Fabry’ego manifestująca się pierwszymi objawami w różnym wieku i z różną intensywnością jest trudna w diagnostyce. W badaniu wykazano, że stosowanie systemów diagnostycznych wspieranych uczeniem maszynowym jako narzędzie przesiewowe jest uzasadnione. Na podstawie uzyskanych wyników

oraz analizie badań nad innymi chorobami rzadkimi np. Amyloidozą, chorobą Gaucher, można stwierdzić, że takie systemy mogą być skuteczne w diagnostyce wielu chorób rzadkich.

W pracy omówiono znaczenie transparentności i interpretowalności modeli w kontekście zrozumienia narzędzi wykorzystujących sztuczną inteligencję w systemach wsparcia decyzji diagnostycznych. W okresie pisania tej pracy transparentność i interpretowalność modeli miała charakter dobrych praktyk w tworzeniu narzędzi AI, jednak nie była regulowana przepisami.

W 2024 Parlament Europejski przyjął akt w sprawie sztucznej inteligencji (AI Act), który ma w pełni obowiązywać w ciągu 24 miesięcy. Systemy wykorzystujące uczenie maszynowe mające związek ze zdrowiem sklasyfikowane są w nim jako systemy wysokiego ryzyka. W artykule 13 AI Act zawarto zalecenie: *“High-risk AI systems shall be designed and developed in such a way as to ensure that their operation is **sufficiently transparent** to enable deployers **to interpret a system’s output** and use it appropriately.”* [82] Tym samym Parlament Europejski w pewnym stopniu nakłada obowiązek tworzenia przejrzystych systemów wsparcia decyzji diagnostycznych działających w oparciu o metody uczenia maszynowego.

7. Bibliografia

- [1] *Artificial Intelligence in Healthcare Market Analysis Report By Component, By Application, By Region, Competitive Insights And Segment Forecasts From 2019 To 2025*. 2020.
- [2] *AI In Healthcare Market Size, Share & Trends Analysis Report By Component (Software Solutions, Hardware, Services), By Application (Virtual Assistants, Connected Machines), By Region, And Segment Forecasts, 2024 - 2030*. 2024.
- [3] Jefferies JL, Spencer AK, Lau HA, et al. A new approach to identifying patients with elevated risk for Fabry disease using a machine learning algorithm. *Orphanet J Rare Dis*; 16. Epub ahead of print 1 December 2021. DOI: 10.1186/s13023-021-02150-3.
- [4] Visibelli A, Roncaglia B, Spiga O, et al. The Impact of Artificial Intelligence in the Odyssey of Rare Diseases. *Biomedicines*; 11. Epub ahead of print 1 March 2023. DOI: 10.3390/biomedicines11030887.
- [5] Bashir AK, Victor N, Bhattacharya S, et al. A Survey on Federated Learning for the Healthcare Metaverse: Concepts, Applications, Challenges, and Future Directions, <http://arxiv.org/abs/2304.00524> (2023).
- [6] Feng Q, Du M, Zou N, et al. Fair Machine Learning in Healthcare: A Review, <http://arxiv.org/abs/2206.14397> (2022).
- [7] Piccialli F, Somma V Di, Giampaolo F, et al. A survey on deep learning in medicine: Why, how and when? *Information Fusion* 2021; 66: 111–137.
- [8] Zhou B, Yang G, Shi Z, et al. Natural Language Processing for Smart Healthcare. Epub ahead of print 18 October 2021. DOI: 10.1109/RBME.2022.3210270.
- [9] Yuan M, Kumar V, Ahmad MA, et al. Assessing Fairness in Classification Parity of Machine Learning Models in Healthcare, <http://arxiv.org/abs/2102.03717> (2021).
- [10] Li G, Xu G, Sangaiah AK, et al. EdgeLaaS: Edge learning as a service for knowledge-centric connected healthcare. *IEEE Netw* 2019; 33: 37–43.
- [11] Chowdhury S, Zhang C, Yu PS, et al. Mixed Pooling Multi-View Attention Autoencoder for Representation Learning in Healthcare, <http://arxiv.org/abs/1910.06456> (2019).
- [12] Chen IY, Joshi S, Ghassemi M, et al. Probabilistic Machine Learning for Healthcare, <http://arxiv.org/abs/2009.11087> (2020).
- [13] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition, <http://arxiv.org/abs/1409.1556> (2014).
- [14] Torfi A, Fox EA. CorGAN: Correlation-Capturing Convolutional Generative Adversarial Networks for Generating Synthetic Healthcare Records, <http://arxiv.org/abs/2001.09346> (2020).
- [15] Vani A, Jernite Y, Sontag D. Grounded Recurrent Neural Networks, <http://arxiv.org/abs/1705.08557> (2017).
- [16] Le H, Tran T, Venkatesh S. Dual Memory Neural Computer for Asynchronous Two-view Sequential Learning, <http://arxiv.org/abs/1802.00662> (2018).

- [17] Taimoor N, Rehman S. Reliable and Resilient AI and IoT-Based Personalised Healthcare Services: A Survey. *IEEE Access* 2022; 10: 535–563.
- [18] Wang Y, Zhao Y, Petzold L. Are Large Language Models Ready for Healthcare? A Comparative Study on Clinical Language Understanding, <http://arxiv.org/abs/2304.05368> (2023).
- [19] Agnihotri P, Ketabi S, Namdar K, et al. *Using Multi-modal Data for Improving Generalizability and Explainability of Disease Classification in Radiology*.
- [20] Jiang F, Jiang Y, Zhi H, et al. Artificial intelligence in healthcare: Past, present and future. *Stroke and Vascular Neurology* 2017; 2: 230–243.
- [21] Wilson A, Chiorean A, Aguiar M, et al. Development of a rare disease algorithm to identify persons at risk of Gaucher disease using electronic health records in the United States. *Orphanet J Rare Dis*; 18. Epub ahead of print 1 December 2023. DOI: 10.1186/s13023-023-02868-2.
- [22] Harrer S, Shah P, Antony B, et al. Artificial Intelligence for Clinical Trial Design. *Trends in Pharmacological Sciences* 2019; 40: 577–591.
- [23] Schwalbe N, Wahl B. Artificial intelligence and the future of global health. *The Lancet* 2020; 395: 1579–1586.
- [24] Poo M ming, Du J lin, Ip NY, et al. China Brain Project: Basic Neuroscience, Brain Diseases, and Brain-Inspired Computing. *Neuron* 2016; 92: 591–596.
- [25] Johnson KB, Wei WQ, Weeraratne D, et al. Precision Medicine, AI, and the Future of Personalized Health Care. *Clinical and Translational Science* 2021; 14: 86–93.
- [26] Osmani V, Li L, Danieletto M, et al. *Processing of Electronic Health Records using Deep Learning: A review*, <https://phekb.org/phenotype/>.
- [27] Tschuchnig ME, Oostingh GJ, Gadermayr M. Generative Adversarial Networks in Digital Pathology: A Survey on Trends and Future Potential. *Patterns*; 1. Epub ahead of print 11 September 2020. DOI: 10.1016/j.patter.2020.100089.
- [28] Dubey S, Tiwari G, Singh S, et al. Using machine learning for healthcare treatment planning. *Front Artif Intell*; 6. Epub ahead of print 2023. DOI: 10.3389/frai.2023.1124182.
- [29] Ahmad A, Tariq A, Hussain HK, et al. BULLET : Jurnal Multidisiplin Ilmu Equity and Artificial Intelligence in Surgical Care: A Comprehensive Review of Current Challenges and Promising Solutions.
- [30] Bouh MM, Hossain F, Ahmed A. A Machine Learning Approach to Digitize Medical History and Archive in a Standard Format. In: *International Conference on Information and Communication Technologies for Ageing Well and e-Health, ICT4AWE - Proceedings*. Science and Technology Publications, Lda, 2023, pp. 230–236.
- [31] Adamson B, Waskom M, Blarre A, et al. Approach to Machine Learning for Extraction of Real-World Data Variables from Electronic Health Records. DOI: 10.1101/2023.03.02.23286522.

- [32] Rider NL, Coffey M, Kurian A, et al. A validated artificial intelligence-based pipeline for population-wide primary immunodeficiency screening. *Journal of Allergy and Clinical Immunology* 2023; 151: 272–279.
- [33] Hoffmann G, Bietenbeck A, Lichtinghagen R, et al. Using machine learning techniques to generate laboratory diagnostic pathways—a case study. *J Lab Precis Med* 2018; 3: 58–58.
- [34] Soltan AAS, Kouchaki S, Zhu T, et al. Rapid triage for COVID-19 using routine clinical data for patients attending hospital: development and prospective validation of an artificial intelligence screening test. *Lancet Digit Health* 2021; 3: e78–e87.
- [35] Taylor AG, Mielke C, Mongan J. Automated detection of moderate and large pneumothorax on frontal chest X-rays using deep convolutional neural networks: A retrospective study. *PLoS Med*; 15. Epub ahead of print 1 November 2018. DOI: 10.1371/journal.pmed.1002697.
- [36] Sullivan BA, Kausch SL, Fairchild KD. Artificial and human intelligence for early identification of neonatal sepsis. *Pediatric Research* 2023; 93: 350–356.
- [37] Addad VV, Palma LMP, Vaisbich MH, et al. A comprehensive model for assessing and classifying patients with thrombotic microangiopathy: the TMA-INSIGHT score. *Thromb J*; 21. Epub ahead of print 1 December 2023. DOI: 10.1186/s12959-023-00564-6.
- [38] Allegra A, Mirabile G, Tonacci A, et al. Machine Learning Approaches in Diagnosis, Prognosis and Treatment Selection of Cardiac Amyloidosis. *International Journal of Molecular Sciences*; 24. Epub ahead of print 1 March 2023. DOI: 10.3390/ijms24065680.
- [39] Kessel K, Mattila J, Linder N, et al. Deep Learning Algorithms for Corneal Amyloid Deposition Quantitation in Familial Amyloidosis. *Ocul Oncol Pathol* 2020; 6: 58–65.
- [40] Martini N, Aimo A, Barison A, et al. Deep learning to diagnose cardiac amyloidosis from cardiovascular magnetic resonance. *Journal of Cardiovascular Magnetic Resonance*; 22. Epub ahead of print 1 December 2020. DOI: 10.1186/s12968-020-00690-4.
- [41] Germain P, Vardazaryan A, Padoy N, et al. Deep Learning Supplants Visual Analysis by Experienced Operators for the Diagnosis of Cardiac Amyloidosis by Cine-CMR. *Diagnostics*; 12. Epub ahead of print 1 January 2022. DOI: 10.3390/diagnostics12010069.
- [42] Reena MR, Ameer PM. Localization and recognition of leukocytes in peripheral blood: A deep learning approach. *Comput Biol Med*; 126. Epub ahead of print 1 November 2020. DOI: 10.1016/j.compbiomed.2020.104034.
- [43] Yang D, Zhao H, Han T, et al. Leukocyte subtypes identification using bilinear self-attention convolutional neural network. *Measurement (Lond)*; 173. Epub ahead of print 1 March 2021. DOI: 10.1016/j.measurement.2020.108643.
- [44] Duan Y, Wang J, Hu M, et al. Leukocyte classification based on spatial and spectral features of microscopic hyperspectral images. *Opt Laser Technol* 2019; 112: 530–538.
- [45] Saeed U, Kumar K, Khuhro MA, et al. DeepLeukNet—A CNN based microscopy adaptation model for acute lymphoblastic leukemia classification. *Multimed Tools Appl* 2024; 83: 21019–21043.

- [46] Oh SH, Back S, Park J. Measuring patient similarity on multiple diseases by joint learning via a convolutional neural network. *Sensors*; 22. Epub ahead of print 1 January 2022. DOI: 10.3390/s22010131.
- [47] Lee GH, Yoon S, Nam M, et al. Performance of digital morphology analyzer CellaVision DC-1. *Clin Chem Lab Med* 2023; 61: 133–141.
- [48] Kouri T, Alagrund K, Lehtonen M, et al. Verification of UriSed 3 PRO automated urine microscope in regional laboratory environment. *Clinica Chimica Acta* 2021; 515: 96–103.
- [49] Luo S, Shi Y, Chin LK, et al. Machine-Learning-Assisted Intelligent Imaging Flow Cytometry: A Review. *Advanced Intelligent Systems*; 3. Epub ahead of print November 2021. DOI: 10.1002/aisy.202100073.
- [50] Ng DP, Zuromski LM. Augmented Human Intelligence and Automated Diagnosis in Flow Cytometry for Hematologic Malignancies. *Am J Clin Pathol* 2021; 155: 597–605.
- [51] Clichet V, Lebon D, Chapuis N, et al. Artificial intelligence to empower diagnosis of myelodysplastic syndromes by multiparametric flow cytometry. *Haematologica* 2023; 108: 2435–2443.
- [52] Sobhani F, Robinson R, Hamidinekoo A, et al. Artificial intelligence and digital pathology: Opportunities and implications for immuno-oncology. *Biochimica et Biophysica Acta - Reviews on Cancer*; 1875. Epub ahead of print 1 April 2021. DOI: 10.1016/j.bbcan.2021.188520.
- [53] Gecer B, Aksoy S, Mercan E, et al. Detection and classification of cancer in whole slide breast histopathology images using deep convolutional networks. *Pattern Recognit* 2018; 84: 345–356.
- [54] Sunnetci KM, Alkan A. Biphasic majority voting-based comparative COVID-19 diagnosis using chest X-ray images. *Expert Systems with Applications*; 216. Epub ahead of print 15 April 2023. DOI: 10.1016/j.eswa.2022.119430.
- [55] Zhang P, Zhong Y, Tang X, et al. *Learning Diagnosis of COVID-19 from a Single Radiological Image*, <https://github.com/PengyiZhang/CoSinGAN>.
- [56] Adams SJ, Henderson RDE, Yi X, et al. Artificial Intelligence Solutions for Analysis of X-ray Images. *Canadian Association of Radiologists Journal* 2021; 72: 60–72.
- [57] Shin H, Park JE, Jun Y, et al. Deep learning referral suggestion and tumour discrimination using explainable artificial intelligence applied to multiparametric MRI. *Eur Radiol* 2023; 33: 5859–5870.
- [58] Dicle O. Artificial intelligence in diagnostic ultrasonography. *Diagnostic and Interventional Radiology* 2023; 29: 40–45.
- [59] Whelehan DF, Conlon KC, Ridgway PF. Medicine and heuristics: cognitive biases and medical decision-making. *Irish Journal of Medical Science (1971 -)* 2020; 189: 1477–1484.
- [60] Yoon JH, Kim EK. Deep learning-based artificial intelligence for mammography. *Korean J Radiol* 2021; 22: 1225–1239.

- [61] Li Q, Li S, Li R, et al. Low-dose computed tomography image reconstruction via a multistage convolutional neural network with autoencoder perceptual loss network. *Quant Imaging Med Surg* 2022; 12: 1929–1957.
- [62] Yamagiwa K, Tsuchiya J, Yokoyama K, et al. Enhancement of 18F-Fluorodeoxyglucose PET Image Quality by Deep-Learning-Based Image Reconstruction Using Advanced Intelligent Clear-IQ Engine in Semiconductor-Based PET/CT Scanners. *Diagnostics*; 12. Epub ahead of print 1 October 2022. DOI: 10.3390/diagnostics12102500.
- [63] Hsieh J, Liu E, Nett B, et al. *A new era of image reconstruction: TrueFidelity™ Technical white paper on deep learning image reconstruction.*
- [64] Singh A, Sengupta S, Lakshminarayanan V. Explainable deep learning models in medical image analysis, <http://arxiv.org/abs/2005.13799> (2020).
- [65] Sharif Rahmani E, Lawarde A, Lingasamy P, et al. MBMethPred: a computational framework for the accurate classification of childhood medulloblastoma subgroups using data integration and AI-based approaches. *Front Genet*; 14. Epub ahead of print 2023. DOI: 10.3389/fgene.2023.1233657.
- [66] Touvron H, Martin L, Stone K, et al. Llama 2: Open Foundation and Fine-Tuned Chat Models, <http://arxiv.org/abs/2307.09288> (2023).
- [67] Brown TB, Mann B, Ryder N, et al. Language Models are Few-Shot Learners, <http://arxiv.org/abs/2005.14165> (2020).
- [68] Zhao WX, Zhou K, Li J, et al. A Survey of Large Language Models, <http://arxiv.org/abs/2303.18223> (2023).
- [69] Singhal K, Azizi S, Tu T, et al. Large language models encode clinical knowledge. *Nature* 2023; 620: 172–180.
- [70] Yang X, Chen A, Pournajatian N, et al. *A Large Language Model for Electronic Health Records*, https://catalog.ngc.nvidia.com/orgs/nvidia/teams/clara/models/gatortron_og.
- [71] Loh HW, Ooi CP, Seoni S, et al. Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011–2022). *Comput Methods Programs Biomed* 2022; 226: 107161.
- [72] Ribeiro MT, Singh S, Guestrin C. ‘Why should i trust you?’ Explaining the predictions of any classifier. In: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Association for Computing Machinery, 2016, pp. 1135–1144.
- [73] Lundberg S, Lee S-I. A Unified Approach to Interpreting Model Predictions, <http://arxiv.org/abs/1705.07874> (2017).
- [74] Gerlings J, Jensen MS, Shollo A. *Explainable AI, but explainable to whom?*
- [75] Haendel M, Vasilevsky N, Unni D, et al. How many rare diseases are there? *Nature Reviews Drug Discovery* 2020; 19: 77–78.
- [76] Willmen T, Völkel L, Ronicke S, et al. Health economic benefits through the use of diagnostic support systems and expert knowledge. *BMC Health Serv Res*; 21. Epub ahead of print 1 December 2021. DOI: 10.1186/s12913-021-06926-y.

- [77] Weinreich SS, Mangon R, Sikkens JJ, et al. [Orphanet: a European database for rare diseases]. *Ned Tijdschr Geneesk* 2008; 152: 518–9.
- [78] Liévin V, Hansen JM, Lund A, et al. FindZebra online search delving into rare disease case reports using natural language processing. *PLOS Digital Health* 2023; 2: e0000269.
- [79] Ronicke S, Hirsch MC, Türk E, et al. Can a decision support system accelerate rare disease diagnosis? Evaluating the potential impact of Ada DX in a retrospective study. *Orphanet J Rare Dis*; 14. Epub ahead of print 21 March 2019. DOI: 10.1186/s13023-019-1040-6.
- [80] Wanner C, Ortiz A, Wilcox WR, et al. Global reach of over 20 years of experience in the patient-centered Fabry Registry: Advancement of Fabry disease expertise and dissemination of real-world evidence to the Fabry community. *Molecular Genetics and Metabolism*; 139. Epub ahead of print 1 July 2023. DOI: 10.1016/j.ymgme.2023.107603.
- [81] Kermond-Marino A, Weng A, Xi Zhang SK, et al. Population Frequency of Undiagnosed Fabry Disease in the General Population. *Kidney Int Rep* 2023; 8: 1373–1379.
- [82] The Artificial Intelligence Act. P9_TA(2024)0138.
https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf.

Streszczenie

W rozprawie doktorskiej przedstawiono potencjał systemów wsparcia decyzji diagnostycznych opartych na uczeniu maszynowym, koncentrując się na dwóch obszarach: wyjaśnialności modeli splotowych sieci neuronowych w klasyfikacji mikroskopowych obrazów leukocytów z rozmazów krwi obwodowej oraz systemu badań przesiewowych w kierunku choroby Andersona-Fabry'ego, wykorzystującego metody przetwarzania języka naturalnego do wyodrębnienia cech klinicznych z nieustrukturyzowanych danych medycznych.

W ramach badań podjęto wyzwanie wyjaśnienia procesu decyzyjnego modeli splotowych sieci neuronowych w klasyfikacji leukocytów na obrazach mikroskopowych z rozmazów krwi obwodowej. Wykorzystując algorytm Shapely Additive Explanations wskazano istotne elementy różnicujące komórki i stwierdzono, że decyzje modeli są zgodne z praktyką laboratoryjną.

Omówiono szanse jakie daje wdrożenie transparentnych systemów wspierających diagnostykę choroby rzadkiej – Andersona-Fabry'ego. Stosując metody przetwarzania języka naturalnego do ekstrakcji cech klinicznych określonych przez lekarzy specjalistów opracowano system przesiewowy dla choroby Andersona-Fabry'ego. System zweryfikowano przesiewając dokumentację medyczną pacjentów jednego z polskich szpitali. Wśród pacjentów z wysokim ryzykiem choroby, lekarze w dalszym postępowaniu diagnostycznym potwierdzili chorobę Andersona-Fabry'ego.

Podsumowując, przeprowadzone badania podkreślają znaczenie sztucznej inteligencji w diagnostyce medycznej. Opracowując zrozumiałe systemy wsparcia decyzji diagnostycznych we współpracy z ekspertami medycznymi, możemy przyspieszyć proces diagnostyczny, tym samym wpłynąć na wyniki leczenia pacjentów. Ponieważ sztuczna inteligencja rozwija się w bardzo szybkim tempie, ważne jest, aby zwracać szczególną uwagę na przejrzystość działania narzędzi, i bezpieczeństwo danych. Zapewniając, że wdrożenie sztucznej inteligencji w systemach wsparcia decyzji diagnostycznych przyniesie korzyści pacjentom i pracownikom ochrony zdrowia.

Summary

The dissertation outlines the potential of machine learning-based diagnostic decision support systems, focusing on two areas: the explainability of spline neural network models in the classification of microscopic images of leukocytes from peripheral blood smears, and an Anderson-Fabry disease screening system using natural language processing methods to extract clinical features from unstructured medical data.

The research addresses the challenge of explaining the decision-making process of spline neural network models in the classification of leukocytes in microscopic images from peripheral blood smears. Using the Shapely Additive Explanations algorithm, important cell differentiation elements were identified and the models' decisions were found to be consistent with laboratory practice.

The opportunities presented by the implementation of transparent systems to support the diagnosis of the rare disease Anderson-Fabry are discussed. Using natural language processing methods to extract clinical features identified by medical specialists, a screening system for Anderson-Fabry disease was developed. The system was validated by screening the medical records of patients from a Polish hospital. Among patients with a high risk of the disease, doctors confirmed Anderson-Fabry disease in further diagnostic procedures.

In conclusion, the study highlights the importance of artificial intelligence in medical diagnosis. By developing comprehensible diagnostic decision support systems in collaboration with medical experts, we can speed up the diagnostic process, thereby influencing patient outcomes. As artificial intelligence is developing at a very fast pace, it is important to pay particular attention to the explainability of the tools, and the security of the data. Ensuring that the implementation of artificial intelligence in diagnostic decision support systems will benefit patients and healthcare professionals.

Życiorys zawodowy

Dane podstawowe:

Adrian Michalski

Adres e-mail: ad.michalski2@gmail.com

Data i miejsce urodzenia: 12.04.1994., Bydgoszcz

Wykształcenie:

10.2018r. – obecnie

Studia doktoranckie z zakresu nauk medycznych, VI rok

10.2013r. - 06.2018r.

Studia stacjonarne jednolite magisterskie na kierunku analityka medyczna, Wydział Farmaceutyczny, Collegium Medicum im. Ludwika Rydygiera w Bydgoszczy, UMK Toruń

Doświadczenie:

04.2023r. – obecnie

Praca na stanowisku Data Scientist w BlueRider.Software Sp. z o.o., ul. Henryka Sienkiewicza 40A, 15-004 Białystok

04.2022r. – 02.2023r.

Praca na stanowisku Lead Data Scientist w Saventic Health Sp. z o.o., ul. Władysława Łokietka 5, 87-100 Toruń, Polska

02.2021r. – 04.2022r.

Praca na stanowisku Data Scientist w Saventic Health Sp. z o.o., ul. Władysława Łokietka 5, 87-100 Toruń, Polska

02.2019r. – 02.2021r.

Praca na stanowisku Przedstawiciela Naukowo-Handlowego w Biomedica Poland sp. z o. o., ul. Raszyńska 13, 05-500 Piaseczno

06.2018r. – 02.2019r.

Praca na stanowisku Diagnosty Laboratoryjnego w Zakładzie Diagnostyki Laboratoryjnej Szpitala Uniwersyteckiego nr 1 im. dr Antoniego Jurasza, ul. Marii Skłodowskiej Curie 9, 85-094 Bydgoszcz

Publikacje:

Michalski A, Olszewska-Stonina D, Hołyńska-Iwan I. Efekt działania alkoholu etylowego na organizm człowieka. W: Nowe spojrzenie na wykorzystanie i aktywność wybranych substancji biologicznie czynnych. P. Kiciński. Lublin: Wydaw. Nauk. TYGIEL, 2018 s. 19-31 (MNIe = 20)

Michalski A, Kupcewicz B. Automatyczna klasyfikacja komórek rozmazu krwi obwodowej na przykładzie zatrucia ołowiem. Farmacja Pol. 2020: T.76. nr 6, S. 318-323 (MNIe = 70)

Kloska A, Tarczewska M, Gielczyk A, Kloska S. M, **Michalski A**, Serafin Z, Woźniak M. Influence of augmentation on the performance of double ResNet-based model for chest X-rays classification. Pol. J. Radiol. 2023 : T.88. nr 1, s. e244-e250. (IF = 1,20 MNIe = 100)

Michalski A, Duraj K, Kupcewicz, B. Leukocyte deep learning classification assessment using Shapley additive explanations algorithm. Int J Lab Hematol. 2023; 1-6. <https://doi.org/10.1111/ijlh.14031> (IF = 3,45 MNIe = 70)

Michalski A, Lis K, Stankiewicz J, Kloska M.S, Dudziński M, Muras-Szwedziak K, Nowicki M, Bazan-Socha S, Dąbrowski J.M, Basak W.G. Supporting the diagnosis of Fabry disease using a natural language processing-based approach. J. Clin. Med. 2023, 12(10), 3599; <https://doi.org/10.3390/jcm12103599> (IF = 4,964 MNIe = 140)

Konferencje i działalność popularyzatorska

Michalski A, Dudziński M. Sztuczna inteligencja: perspektywy w amyloidozie i innych chorobach rzadkich. Amyloidoza V Interdyscyplinarna Konferencja. Warszawa 8-9.10.2021

Michalski A, Lis K. Sztuczna inteligencja w hematologii. Fundacja Saventic. Warszawa 23.11.2021

Michalski A, Stankiewicz J. Diagnostyka pierwotnych niedoborów odporności z wykorzystaniem uczenia maszynowego. Fundacja Saventic. Warszawa, 06.05.2022

Michalski A, Dudziński M. Czym jest AI – Diagnostyka choroby Castlemana. XXX Zjazd PTHiT. Bydgoszcz 8-10.09.2022

Michalski A. Data Science w ochronie zdrowia – jak możemy pomóc pacjentom. PyData Bydgoszcz. Bydgoszcz 03.11.2022

Michalski A, Kottan S. Wykorzystanie sztucznej inteligencji do wczesnego wykrywania wrodzonych błędów odporności w POZ – teraźniejszość i przyszłość. III Bydgoska Jesień Immunologiczna. Bydgoszcz 4-5.11.2022

Michalski A. Wizualizacja procesów i idei. Data Science Summit. Warszawa 23-24.2023

Michalski A. AI w ochronie zdrowia. Horyzont.AI. Bydgoszcz. 05.02.2024

Granty

Współudział w grantie zewnętrznym.

„Opracowywanie i wdrożenie innowacyjnego, opartego o uczenie maszynowe procesu diagnostycznego, w celu skrócenia ścieżki pacjenta i poprawy jakości opieki zdrowotnej nad pacjentami z chorobami limfo- i mieloproliferacyjnymi” realizowanego w ramach Programu Operacyjnego Inteligentny Rozwój. Nr wniosku: POIR.01.01.01-00-0688/20

Załącznik I

ORIGINAL ARTICLE

Leukocyte deep learning classification assessment using Shapley additive explanations algorithm

Adrian Michalski¹ | Konrad Duraj² | Bogumiła Kupcewicz¹

¹Department of Analytical Chemistry, Faculty of Pharmacy, University of Nicolaus Copernicus, Collegium Medicum, Bydgoszcz, Poland

²Department of Biosensors and Processing of Biomedical Signals, Faculty of Biomedical Engineering, Silesian University of Technology, Zabrze, Poland

Correspondence

Adrian Michalski, Department of Analytical Chemistry, Faculty of Pharmacy, University of Nicolaus Copernicus, Collegium Medicum, Jagiellońska 15, 85-067 Bydgoszcz, Poland.
Email: ad.michalski2@gmail.com

Abstract

Introduction: A peripheral blood smear is a basic test for hematological disease diagnosis. This test is performed manually in many places worldwide, which requires both time and qualified staff. Large laboratories are equipped with digital morphology analyzers, some of which are based on deep learning methods. However, it is difficult to explain to scientists how they work. In this paper, we proposed to add an explanatory factor to enhance the interpretability of deep learning models in leukocyte classification.

Methods: 10 297 single images of leukocytes obtained from peripheral blood smears were included in this study. Pre-trained and fully trained VGG16 and VGG19 models were used to classify the leukocytes, and Shapley Additive Explanations (SHAP) DeepExplainer was applied to visualize the area of cells that were significant for classification. The output images from the DeepExplainer were compared with cellular elements that are essential to laboratory practice.

Results: The accuracy of our fully trained models was 99.81% for VGG16 and 99.79% for VGG19. It achieved slightly better results than the partially trained model, which scored 98.67% for VGG16 and 98.33% for VGG19. Their SHAP explanations indicated the significance of cellular structures in microscopic examination. Explanations in the pre-trained models have proved the cell and nucleus contours to be relevant to classification, while explanations in the fully trained models pointed to the cytoplasm area.

Conclusion: Despite different SHAP DeepExplainer explanations for fully and partially trained models, this method appears to be helpful for the verification of leukocyte classification in automated peripheral blood smear examination.

KEYWORDS

blood smear, deep learning, leukocyte classification, model explanation, SHAP explainer

1 | INTRODUCTION

Cells of all hematopoietic lines, except the red blood cells (RBC), are involved in the functioning of the immune system. The proper functioning of the immune system is ensured, among other things, by a normal percentage of leukocytes, also known as white blood cells. Mature leukocytes in the blood are divided into five types:

neutrophils, eosinophils, basophils (granulocytes), lymphocytes, and monocytes (agranulocytes).¹ A test that evaluates the number and type of cells circulating in the blood is a complete blood count with leukocyte differential (CBC DIFF). Any flagged abnormalities in this test indicate that scientists should further investigate the sample by performing a blood smear examination.² This test evaluates the morphology of platelets, red blood cells, and white blood cells, as well as

differentiates them. In most laboratories, scientists still manually stain films and perform blood smear examinations on optical microscopes. They distinguish leukocytes by their features: cell size and shape, nuclei shape, presence and color of granularity.³

To reduce the likelihood of error and shorten testing times, digital morphology analyzers have been developed, including those using artificial intelligence classification methods. However, automated solutions are much more expensive than manual testing and require a broader explanation for scientists unfamiliar with deep learning methods.⁴ The accuracy of these analyzers is similar to manual testing for five normal leukocytes; however, blood smears with pathological cells should be confirmed by microscopic examination.⁵ In the past, many researchers achieved satisfying results in leukocyte classification with machine learning approaches. They tested methods such as support vector machine (SVM),⁶ Naive Bayes classifiers,⁷ and learning discriminant analysis (LDA).⁸ Nowadays, neural networks, mainly convolutional neural networks (CNN), are harnessed to solve the problem. Scientists who applied these methods achieved a high accuracy of 95%–99.8%, also using the transfer learning approach.^{4,9,10}

In comparison with the previously used image classification algorithms, the main advantages of CNN are less demanding image pre-processing and no need to indicate essential features for classification.¹¹ CNN can be trained fully or partially using a pre-trained model, which is the base of the transfer learning approach. Applying a pre-trained model is possible by calculating the weights in the network layers to generalize the image classification problem using features such as shapes, edges, and pixel intensities. By fine-tuning such models, we can reach faster convergence.¹² Despite the good effectiveness of CNN models, scientists may question the explainability of the classification model. Due to the multilayer nonlinear structure, neural networks are considered black boxes. Therefore, the calculations behind the CNN models' predictions are not traceable by humans.

Several black box explanatory tools have been developed over the past few years.¹³ One of them is SHAP DeepExplainer—an approach to explain any machine learning model. It uses the Shapley value originating in game theory to determine the validity of the output features of the model.¹⁴ The Shapley value was developed to determine a player's contribution to the final outcome of the games. Based on this value, the salary was to be shared fairly among players. The Shapley value reasonably breaks down the features considering the model's output. The importance of a feature is determined by its value and direction - the answer to whether the feature brings us closer to winning the game.

The purpose of this study was to show that pre-trained leukocyte classification models enhanced with explanatory tools, such as SHAP DeepExplainer, could have applications in laboratory diagnostics. We attempted to explain how leukocyte classification works using SHAP, and indicate which cellular elements are most relevant in distinguishing leukocytes. Our results were based on the models we trained, namely convolutional neural network models VGG16 and VGG19.¹⁵

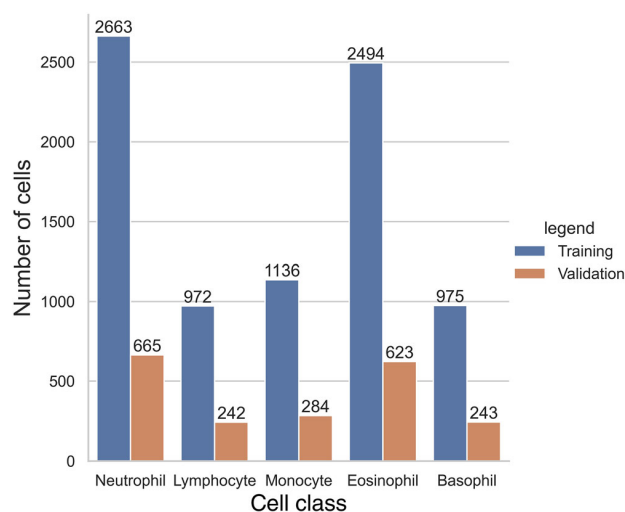


FIGURE 1 Class distribution in train and validation data set

2 | MATERIALS AND METHODS

We conducted a study on a data set provided by Acevedo et al. The collection of peripheral blood images taken on the CellaVision DM96 analyzer contains 17 092 jpg format images of single cells with a size of 360 x 363 pixels and RGB color space. Clinicians from the Hospital Clinic labeled the images. The data set includes eight classes.¹⁶ We focused on neutrophils, lymphocytes, monocytes, eosinophils, and basophils—all of them are found in normal blood smears.¹ We used two convolutional neural networks, VGG16 and VGG19, with a transfer learning approach. SHAP DeepExplainer has been applied to find significant parts of images that the model used to classify leukocytes.

2.1 | Data preparation

The data set includes 10 297 single-cell of five types of leukocyte images with a size of 360 x 363 pixels. We used the original size of images with the channel last format. Images of the cells have been divided into training and validation sets, with 8240 and 2057 samples, respectively. The number of images in each class of leukocytes is shown in Figure 1.

2.2 | Transfer learning

Training deep neural networks requires a large data set and a high processing power, but a sufficient data set is rarely available. The solution to this problem is the transfer learning approach. It uses the core of a previously trained network with its already calculated weights. To solve a specific issue, the network needs to be fine-tuned. When applying transfer learning, it is crucial to choose the optimal strategy for training based on the problem setting (similar or different) and

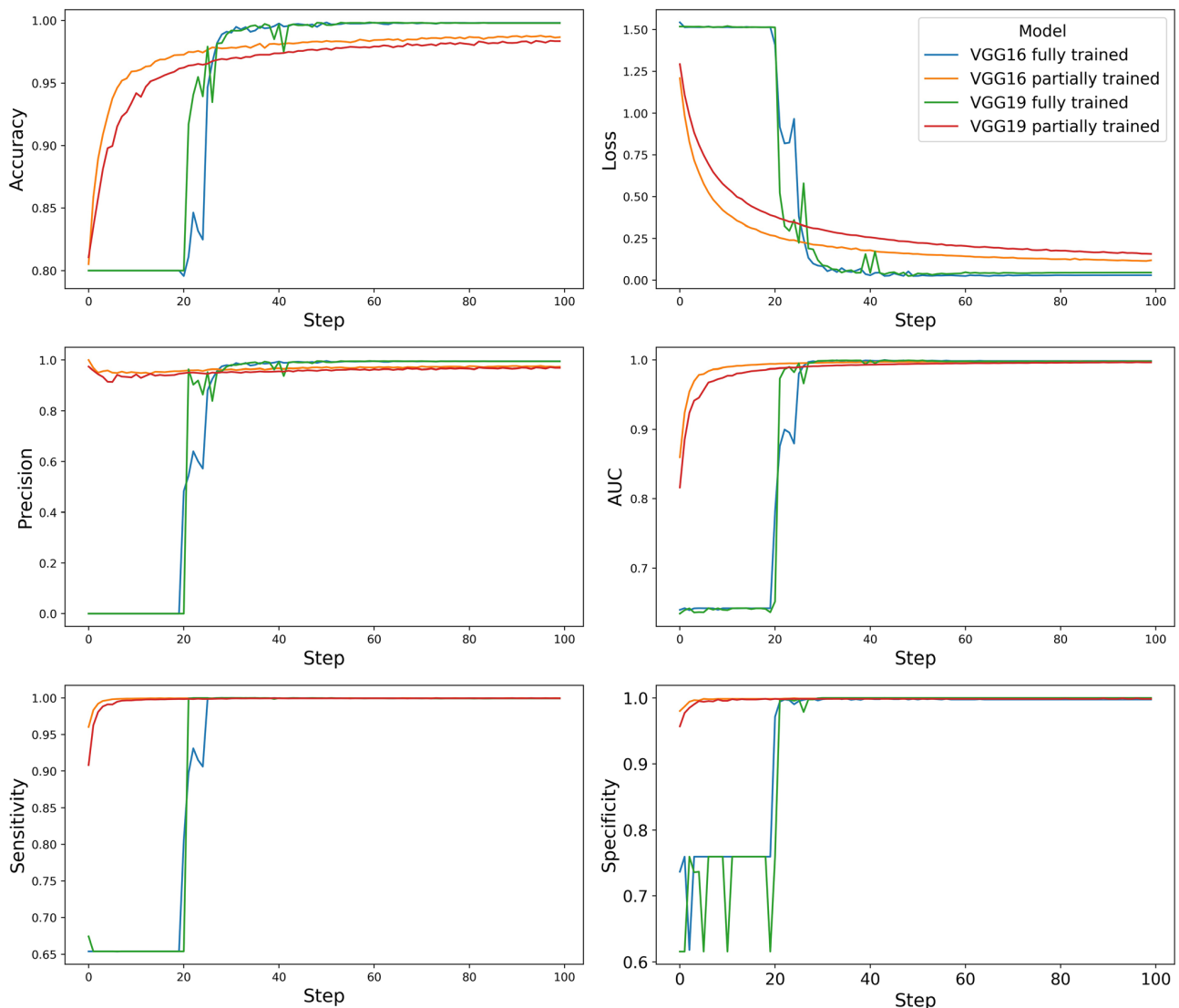


FIGURE 2 Metrics of VGG16 and VGG19 models on validation data set with two schemes—fully trained and partially trained

available data (small or large data set).^{17,18} Given that the data domain for the target data set is different from the source data set (ImageNet¹⁹) and consists of more than a few hundred samples, the problem was categorized as a different and large data set setting. In that setting, it is recommended not to freeze the feature-extractor weights fully. In this paper, we have decided to take two approaches. The first one involved freezing the feature-extractor layers and only fine-tuning the last layer for the classification task. The second one focuses on fine-tuning the entire network. We then compare the trained classifiers in their numerical metrics, computation resources, and SHAP explainability results. Two popular deep learning architectures were used for this task—VGG16 and VGG19. Both networks consist of sequentially stacked convolutional, max-pooling, and batch normalization blocks. All models were trained using Adam optimizer²⁰ with a decaying learning rate (starting from 0.001) and batch size of 32.

2.3 | SHAP

Leukocyte recognition and classification tasks are broadly described, but no attempt has been made to describe the relevant features for leukocyte classification models.^{9,10} In this work, we used SHAP Deep Explainer to describe these features. DeepExplainer is an approximation algorithm for SHAP values in deep learning models built on DeepLIFT.²¹ DeepExplainer, unlike the original DeepLIFT, uses a background sample distribution instead of a single reference value and Shapley's equations to linearize components such as softmax, max, and divisions. The output from Deep Explainer is an image with red and blue pixels. Red pixels are significant positive elements for classification, while blue pixels are significant negative elements. We used 10 random images (two images each class) as a background for all models. Finally, we compared the original images and SHAP output. The logic behind the comparison process was to extract the cell and

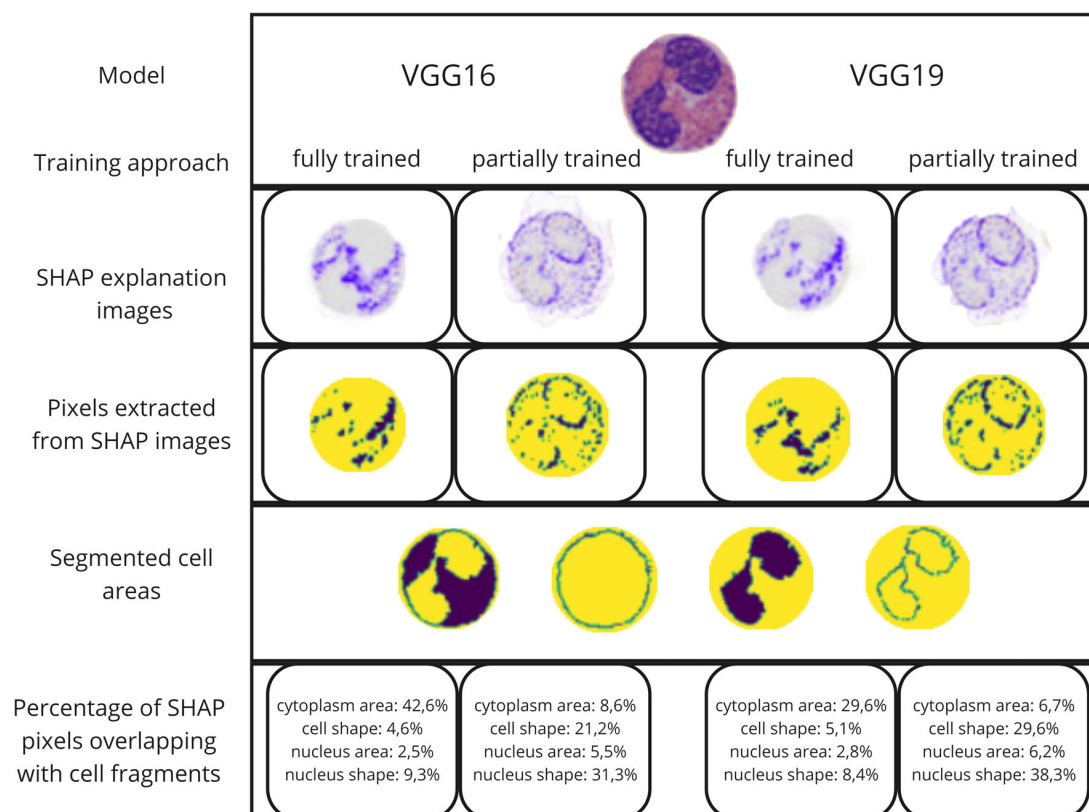


FIGURE 3 Scheme for segmentation and comparison of eosinophil elements and SHAP output

nucleus contour, the cytoplasm and nucleus area and then calculate the percentage of pixels in the aforementioned elements of the cell. If the percentage of shared pixels in a cellular element was more than 30% of all pixels in the SHAP image, we counted this element as significant for classification. 30% cut-off was set arbitrarily.

3 | RESULTS

We have trained two transfer learning models using two approaches: fully trained and partially trained VGG16 and VGG19. Although both training strategies provided sufficient results, it is worth noting that fully-trained models were more precise in their predictions (Figure 2). Multiple test iterations generated similar results. Using the transfer learning method, partially trained models achieved high efficiency faster than their fully trained counterparts (Figure 2). All models were used for classification and SHAP explanation. Then, we extracted the indicated cellular elements by comparing SHAP images with the original ones (Figure 3). In partially trained models, basophils were primarily identified by cell contour and cytoplasm area. On the other hand, eosinophils, neutrophils, and monocytes were identified by cell and nucleus contour. Lymphocytes were classified mainly by cell contour and nucleus area. Whereas in fully trained models, cytoplasm area was relevant for the classification of all cells. The nucleus area was also important

for monocyte classification, while for neutrophils it was the nucleus contour (Table 1).

4 | DISCUSSION

This paper considers the explainability of pre-trained leukocyte classification models using the SHAP method. We demonstrated that models based on VGG16 and VGG19 convolutional neural networks trained fully and partially with pre-trained weights achieve high efficiency. It was also shown that, although comparable results can be achieved in a shorter training time using a frozen feature-extractor scheme with added trainable layers on top, valuable information from the model interpretability standpoint may be lost. As long as the right data set is available, we should train all network parameters. Furthermore, the learning curves indicate that the convergence of the fully trained network needs a lower learning rate (Figure 2). In comparison, the feature-extractor-freeze scheme can handle the larger learning rate, possibly due to the much lower number of parameters to optimize.

Better explainability of the models helps specialists understand the algorithm behind its decision. This seems especially important for deep learning methods used in medical diagnostic systems. Therefore, the motivation of this study was to show that the SHAP algorithm can be applied to explain cell classification in digital morphology analyzers based on such methods. We illustrated the results of the SHAP

TABLE 1 Number of leukocytes classified based on the extracted image elements in the validation data set

Classified leukocytes	Leukocyte elements	VGG16		VGG19	
		Fully trained	Partially trained	Fully trained	Partially trained
		n (%)			
Neutrophils (n = 665)	Cell contour	1 (0.15)	502 (75.49) ^a	0 (0.00)	558 (83.91) ^a
	Nucleus contour	270 (40.60) ^a	617 (92.78) ^a	189 (28.42) ^a	632 (95.04) ^a
	Cytoplasm area	622 (93.53) ^a	47 (7.07)	635 (95.49) ^a	25 (3.76)
	Nucleus area	10 (1.50)	22 (3.31)	13 (1.95)	7 (1.05)
Lymphocytes (n = 242)	Cell contour	78 (32.23) ^a	141 (58.26) ^a	42 (17.36) ^a	156 (64.46) ^a
	Nucleus contour	3 (1.24)	37 (15.29)	7 (2.89)	12 (4.96)
	Cytoplasm area	90 (37.19)	41 (16.94)	99 (40.90)	29 (11.98)
	Nucleus area	123 (50.82) ^a	107 (44.21) ^a	108 (44.62) ^a	82 (33.88) ^a
Monocytes (n = 284)	Cell contour	0 (0.00)	209 (73.59) ^a	0 (0.00)	180 (63.38) ^a
	Nucleus contour	4 (1.41)	146 (51.41) ^a	1 (0.35)	136 (47.89) ^a
	Cytoplasm area	204 (71.83) ^a	41 (14.44)	176 (61.97) ^a	27 (9.51)
	Nucleus area	99 (34.86) ^a	16 (5.63)	132 (46.48) ^a	9 (3.17)
Eosinophils (n = 623)	Cell contour	5 (0.80)	500 (80.26) ^a	7 (1.12)	589 (94.54) ^a
	Nucleus contour	87 (13.96) ^a	422 (67.74) ^a	122 (19.58) ^a	465 (74.64) ^a
	Cytoplasm area	536 (86.04) ^a	6 (0.96)	520 (83.47) ^a	8 (1.28)
	Nucleus area	1 (0.16)	24 (3.85)	2 (0.32)	46 (7.38)
Basophils (n = 243)	Cell contour	2 (0.82)	209 (86.01) ^a	9 (3.70)	218 (89.71) ^a
	Nucleus contour	15 (6.17)	4 (1.65)	9 (3.70)	7 (2.88)
	Cytoplasm area	204 (83.95) ^a	101 (41.56) ^a	221 (90.95) ^a	88 (36.21) ^a
	Nucleus area	67 (27.57) ^a	3 (1.23%)	31 (12.76) ^a	0 (0.00)

^aThe two most frequent combinations of relevant cellular elements for classification.

DeepExplainer, taking into consideration which cellular elements are indicated by the algorithm. We extracted four elements in single cell images and checked whether the pixels from the SHAP output images overlapped with them. It is worth noting that the SHAP explanation does not allow us to say whether the model classified the cell based on its diameter, the ratio of nucleus to cytoplasm or the color of the granules. The SHAP explanation allows us to conclude which area of the image was relevant for cell classification e.g. the area that overlaps with the cell nucleus or the cell contour.

Using the partially trained model, the nucleus and cell contour were relevant for the classification of eosinophils, neutrophils and monocytes. The cell contour was also a distinguishing feature of basophils and lymphocytes, while the contour of the nucleus was not. The reason, for basophils, might be the overlapping of large granules on the cell nucleus, as both cell elements are similarly stained. For lymphocytes, it might be almost the same nucleus contour and cell contour. Instead of the nucleus contour, the cytoplasmic area was associated with the classification of 36.21% (VGG16) and 41.56% (VGG19) of basophils confirming the above conjecture. The cytoplasmic region was decisive for classifying all cells in fully trained models. However, the main lymphocyte classification feature of 50.82% (VGG16) and 44.62% (VGG19) was the nucleus area. The SHAP authors emphasize that not only the presence of certain elements in

the image is important for cell classification, but also its absence. This is supported by the high percentage of leukocytes classified based on cytoplasmic region, in the case of the fully trained model. Since the cytoplasm region was important for both agranulocytes and granulocytes, we can assume that the presence of granules, their color and shape were important for classification. Fully trained model will not be useful, for example, to distinguish between band and segmented neutrophils, which can be distinguished by the nucleus shape. We have identified the most frequent combinations for each cellular element. We found that for neutrophils, monocytes and basophils, they correspond to the elements indicated as the two most relevant for classification. In classification by fully trained model, the combination of cell contour with nucleus area was more frequent for lymphocytes than cytoplasm area with nucleus area. This means that despite the lower number of lymphocytes classified by cytoplasm area, this cellular element may be a more important feature in classification than nucleus area.

5 | CONCLUSIONS

This study explained leukocyte single image classification based on deep learning methods. We proposed VGG16 and VGG19 models

trained with two variations—fully trained and partially trained. Both models achieve a high efficiency of over 98%. However, they vary in the SHAP DeepExplainer explanation. Differentiation of leukocytes by partially trained models is mainly based on cell and nucleus contours, whereas fully trained models' decisions are based on cytoplasmic area.

SHAP DeepExplainer could be used in the development of diagnostic systems such as automatic blood smear examinations. This would increase laboratory analyst's understanding of analyzer performance and allow verification of the correctness of the cell classification.

AUTHOR CONTRIBUTIONS

Adrian Michalski was responsible for the design of the study, implementation of the computer code and supporting algorithms, and writing the manuscript. Konrad Duraj was responsible for the implementation of the computer code and supporting algorithms, and revising the manuscript. Bogumiła Kupcewicz was responsible for the design of the study, and for revising the manuscript. All authors reviewed and approved the final manuscript.

ACKNOWLEDGMENT

The authors express special thanks to Marta Osirńska for verifying the linguistic correctness of this manuscript.

CONFLICT OF INTEREST STATEMENT


The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in Mendeley Data at <https://data.mendeley.com/datasets/snk93bnjr/1>, reference number 10.17632/snk93bnjr.1.

ORCID

Adrian Michalski  <https://orcid.org/0000-0003-0800-5394>

Bogumiła Kupcewicz  <https://orcid.org/0000-0002-4480-7338>

REFERENCES

- Jacobsen SEW, Nerlov C. Haematopoiesis in the era of advanced single-cell technologies. *Nat Cell Biol.* 2019;21(1):2-8. doi:10.1038/s41556-018-0227-8
- Leung E, Johnston A, Olsen B, et al. Laboratory practices for manual blood film review: results of an IQMH patterns of practice survey. *Int J Lab Hematol.* 2021;43(2):184-190. doi:10.1111/ijlh.13343
- Hegde RB, Prasad K, Hebbar H, Sandhya I. Peripheral blood smear analysis using image processing approach for diagnostic purposes: a review. *Biocybern Biomed Eng.* 2018;38(3):467-480. doi:10.1016/j.bbe.2018.03.002
- Shahin AI, Guo Y, Amin KM, Sharawi AA. White blood cells identification system based on convolutional deep neural learning networks. *Comput Methods Programs Biomed.* 2019;168:69-80. doi:10.1016/j.cmpb.2017.11.015
- Kratz A, Lee S h, Zini G, Riedl JA, Hur M, Machin S. Digital morphology analyzers in hematology: ICSH review and recommendations. *Int J Lab Hematol.* 2019;41(4):437-447. doi:10.1111/ijlh.13042
- Duan Y, Wang J, Hu M, et al. Leukocyte classification based on spatial and spectral features of microscopic hyperspectral images. *Opt Laser Technol.* 2019;112:530-538. doi:10.1016/j.optlastec.2018.11.057
- IEEE Singapore Section, IEEE Region 10, Institute of Electrical and Electronics Engineers. *Proceedings of the 2016 IEEE Region 10 Conference (TENCON).* 2016.
- Molina A, Alferez S, Boldu L, Acevedo A, Rodellar J, Merino A. Sequential classification system for recognition of malaria infection using peripheral blood cell images. *J Clin Pathol.* 2020;73(10):665-670. doi:10.1136/jclinpath-2019-206419
- Reena MR, Ameer PM. Localization and recognition of leukocytes in peripheral blood: a deep learning approach. *Comput Biol Med.* 2020;126:126. doi:10.1016/j.compbimed.2020.104034
- Yang D, Zhao H, Han T, Kang Q, Ma J, Lu H. Leukocyte subtypes identification using bilinear self-attention convolutional neural network. *Measurement (Lond).* 2021;173:173. doi:10.1016/j.measurement.2020.108643
- Piccialli F, di Somma V, Giampaolo F, Cuomo S, Fortino G. A survey on deep learning in medicine: why, how and when? *Information Fusion.* 2021;66:111-137. doi:10.1016/j.inffus.2020.09.006
- Zhuang F, Qi Z, Duan K, et al. A comprehensive survey on transfer learning. *Proc IEEE.* 2021;109(1):43-76. doi:10.1109/JPROC.2020.3004555
- Singh A, Sengupta S, Lakshminarayanan V. Explainable deep learning models in medical image analysis. 2020. <http://arxiv.org/abs/2005.13799>.
- Lundberg S, Lee SI. A unified approach to interpreting model predictions. 2017. <http://arxiv.org/abs/1705.07874>.
- Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 2014. <http://arxiv.org/abs/1409.1556>
- Acevedo A, Merino A, Alferez S, Molina A, Boldu L, Rodellar J. A dataset of microscopic peripheral blood cell images for development of automatic recognition systems. *Data Brief.* 2020;30:105474. doi:10.1016/j.dib.2020.105474
- Tan C, Sun F, Kong T, Zhang W, Yang C, Liu C. A survey on deep transfer learning. 2018. <http://arxiv.org/abs/1808.01974>.
- Peng X, Huang Z, Sun X, Saenko K. Domain agnostic learning with disentangled representations. 2019. <http://arxiv.org/abs/1904.12347>.
- Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. *IEEE Conference on Computer Vision and Pattern Recognition, 2009, CVPR 2009.* IEEE; 2009.
- Kingma DP, Ba J. Adam: a method for stochastic optimization. *CoRR, abs/1412.6980.* 2014. <http://arxiv.org/abs/1412.6980>.
- Shrikumar A, Greenside P, Kundaje A. Learning important features through propagating activation differences. *Proceedings of the 34th ICML.* 2017;70:3145-3153. <http://arxiv.org/abs/1704.02685>

How to cite this article: Michalski A, Duraj K, Kupcewicz B. Leukocyte deep learning classification assessment using Shapley additive explanations algorithm. *Int J Lab Hematol.* 2023;1-6. doi:10.1111/ijlh.14031

Załącznik II



Article

Supporting the Diagnosis of Fabry Disease Using a Natural Language Processing-Based Approach

Adrian A. Michalski ^{1,2,†} , Karol Lis ^{1,3,*} , Joanna Stankiewicz ^{1,4}, Sylwester M. Kloska ^{1,5}, Arkadiusz Sycz ^{1,6} , Marek Dudziński ^{1,7}, Katarzyna Muras-Szwedziak ^{8,9} , Michał Nowicki ^{8,9} , Stanisława Bazan-Socha ^{8,10} , Michał J. Dabrowski ^{1,11,‡} and Grzegorz W. Basak ^{1,3,‡}

- ¹ Saventic Health, Polna 66/12 Street, 87-100 Torun, Poland; ad.michalski2@gmail.com (A.A.M.); michal.kloska@saventic.com (S.M.K.); arkadiusz.sycz@saventic.com (A.S.)
- ² Department of Analytical Chemistry, Nicolaus Copernicus University Ludwik Rydygier Collegium Medicum, 85-089 Bydgoszcz, Poland
- ³ Department of Hematology, Transplantation and Internal Medicine, Medical University of Warsaw, 02-097 Warsaw, Poland
- ⁴ Department of Pediatrics, Hematology and Oncology, Nicolaus Copernicus University Ludwik Rydygier Collegium Medicum, 85-094 Bydgoszcz, Poland
- ⁵ Department of Forensic Medicine, Nicolaus Copernicus University Ludwik Rydygier Collegium Medicum, 85-067 Bydgoszcz, Poland
- ⁶ Faculty of Mathematics and Information Science, Warsaw University of Technology, 00-662 Warsaw, Poland
- ⁷ Department of Hematology, Institute of Medical Sciences, College of Medical Sciences, University of Rzeszow, 35-959 Rzeszow, Poland
- ⁸ Saventic Foundation, Polna 66/12 Street, 87-100 Torun, Poland
- ⁹ Department of Nephrology, Hypertension and Kidney Transplantation, Medical University of Lodz, 90-419 Lodz, Poland
- ¹⁰ Department of Internal Medicine, Faculty of Medicine, Jagiellonian University Medical College, 31-008 Krakow, Poland
- ¹¹ Computational Biology Group, Institute of Computer Science of the Polish Academy of Sciences, 01-248 Warsaw, Poland

* Correspondence: karol.lis.publikacje@saventic.com

† These authors contributed equally to this work.

‡ These authors contributed equally to this work.



Citation: Michalski, A.A.; Lis, K.; Stankiewicz, J.; Kloska, S.M.; Sycz, A.; Dudziński, M.; Muras-Szwedziak, K.; Nowicki, M.; Bazan-Socha, S.; Dabrowski, M.J.; et al. Supporting the Diagnosis of Fabry Disease Using a Natural Language Processing-Based Approach. *J. Clin. Med.* **2023**, *12*, 3599. <https://doi.org/10.3390/jcm12103599>

Academic Editor: Álvaro Hermida-Ameijeiras

Received: 10 February 2023

Revised: 1 May 2023

Accepted: 15 May 2023

Published: 22 May 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: In clinical practice, the consideration of non-specific symptoms of rare diseases in order to make a correct and timely diagnosis is often challenging. To support physicians, we developed a decision-support scoring system on the basis of retrospective research. Based on the literature and expert knowledge, we identified clinical features typical for Fabry disease (FD). Natural language processing (NLP) was used to evaluate patients' electronic health records (EHRs) to obtain detailed information about FD-specific patient characteristics. The NLP-determined elements, laboratory test results, and ICD-10 codes were transformed and grouped into pre-defined FD-specific clinical features that were scored in the context of their significance in the FD signs. The sum of clinical feature scores constituted the FD risk score. Then, medical records of patients with the highest FD risk score were reviewed by physicians who decided whether to refer a patient for additional tests or not. One patient who obtained a high-FD risk score was referred for DBS assay and confirmed to have FD. The presented NLP-based, decision-support scoring system achieved AUC of 0.998, which demonstrates that the applied approach enables for accurate identification of FD-suspected patients, with a high discrimination power.

Keywords: clinical diagnosis support system; decision-support; electronic health record; EHR; Fabry disease; natural language processing; NLP; rare disease; risk factor

1. Introduction

Fabry disease (FD) is an ultra-rare genetic disorder with an incidence of 1:40,000 to 1:117,000, depending on the studied population [1]. It is caused by pathological variants of the alpha-galactosidase A gene (*GLA*) encoding lysosomal enzyme alpha-galactosidase A [2]. Lower activity of that enzyme leads to pathological accumulation of glycosphingolipids, such as globotriaosylceramides inside lysosomes, followed by cell degeneration, tissue inflammation, and organ impairment [3]. Since FD is an X-linked disorder, the symptoms are usually more severe in males than in females [4,5]. Interestingly, screening studies have revealed that the frequency of *GLA* pathological variants in newborns is much higher than the prevalence of clinically significant FD cases [6,7]. This phenomenon is likely a result of the complex gene expression profile or post-transcriptional modifications opposing the clinical role of the genetic *GLA* abnormalities [8–10]. Usually, the first clinical symptoms of FD appear early, during the preschool or school periods [11]. They include neuropathic pain (e.g., acroparesthesias), with the so-called “pain crises”, gastrointestinal symptoms, impaired sweating, and cold/heat intolerance. Other characteristic FD traits include recurrent fever, typical skin changes (e.g., angiokeratoma), ophthalmologic (e.g., cornea verticillata) and otolaryngological (e.g., hearing impairment or loss, dizziness) abnormalities, while those most frequently responsible for shortening the patient’s life include cardiac dysfunction, cardiovascular and cerebrovascular events, and chronic kidney disease, usually with proteinuria [12,13]. Despite many symptoms, diagnosis of FD in relation to the appearance of the first symptoms is usually significantly delayed (by 7–10 years), because most of the symptoms are non-specific [14–16]. Furthermore, awareness about this disease in the medical community is low [16,17]. It has been shown that 43% of pediatricians and 70% of rheumatologists could not point to the proper diagnostic approach in FD suspicion [16]. On the other hand, since effective therapy exists [18], early diagnosis is critical to limit disease progression; however, damage done to the organs prior to its initiation cannot be reversed [19–21]. Automatization of the screening process, based on approaches evaluating medical file records for suspicion of FD trait, might facilitate and shorten the time to diagnosis, benefiting patients and their physicians. Several applications have been described in the literature based on artificial intelligence (AI) or related methods for identifying rare diseases. For example, Hughes et al. [22] presented an approach helping to determine the severity of FD using an expert algorithm based on symptoms and clinical findings. Jefferies et al. [23] described an AI tool (OM1 Patient Finder™, OM1 Inc., Boston, MA, USA) capable of identifying patients at high FD risk. They proved the tool to be effective, with an AUC of 0.82. Access to large-scale databases of EHRs presents new opportunities to apply NLP methods and new sources of clinically valuable knowledge. One of the text processing tasks is the recognition of entities, especially in the medicine domain entities like symptoms or signs of certain conditions in general. Today’s implementations of NLP on unstructured narrative texts from EHRs have shown promising results on this task [24]. There is an increasing interest in BERT-like transformers [25,26]. Rule-based approaches have also been shown to be effective [24,27–29]. However, we decided to use a lexicon-based approach similar to that used by Oladapo Oyeboode [29], but instead of the sentiment score, we used sentence embedding, implemented in spaCy [30], and cosine similarity to measure the level of similarity to achieve a complete automated process of extracting data from electronic health records (EHRs). It is important to note that medical description accounts for a substantial part of EHRs, and in many cases the majority of the information is in the form of unstructured data; therefore, NLP-based methods may be more sensitive than approaches based on ICD-9- or ICD-10-derived clinical phenotypes or on laboratory test results alone. Our study aimed to develop and test a risk-factor-based scoring system to support physicians in early diagnosis of FD through automatic screening and analysis of EHRs available in the information systems of primary care and outpatient clinics in real life with the use of NLP. Specific rare disease indication is much more efficient when a patient’s description is mapped against thousands of other records. In addition, tremendous amounts of complicated patient background information—i.e., the complete

documentation of the patient's medical history, which facilitates diagnosis—cannot be evaluated by a physician during one visit, in contrast to the presented approach. We believe that the NLP-based approach presented in this paper could accelerate accurate diagnosis among patients with signs of FD.

2. Materials and Methods

2.1. Studied Population

The study was conducted on medical data obtained from the Saventic Health database that covered eight Polish hospitals that had signed agreements with Saventic Health. It covers years 2008–2022, and all medical information of the patients were included in the study. However, in terms of the time horizon of the medical history, 25% of patients had less than 1 year of historical documentation, 50% less than 5 years, and 75% less than 9 years, respectively. The longest time range for a patient was 16 years between the first and last visit. This database was based on the PostgreSQL Relational Database and contained medical records that had been anonymized in accordance with the GDPR guidelines. The following types of data in unstructured form were present in the patients' electronic health records (EHRs): age and gender, medical history, symptoms and signs, laboratory results, diagnoses, ICD-10 codes, imaging study descriptions, and epicrisis. Patient data were extracted and processed in Jupyter Notebook [31] in Python [32]. The Numpy [33] and Pandas [34] libraries were used for data preprocessing. The informed consent of the patients was not needed to perform analysis based on medical data from health care centers. Data were anonymized by the medical center prior to sharing data with the Saventic. After obtaining the results of the analysis (see Section 3), medical coordinators in individual facilities were contacted and the report specifying high-risk medical records was shared with them. The decision on whether to invite patients for further consultations, including a DBS, was taken by doctors leading individual patients. At first, it was verified, based on the presence of the ICD-10 code E75.2 in the EHRs, that 13 of the patients in the medical database had previously been diagnosed with FD using dry blood spot (DBS) assay. The EHRs of these 13 patients were obtained from the seven Polish hospitals at which they were treated. Together, these patients formed the study group. The control group was established with the aim of imitating a hospital population, in which this kind of scoring system could be implemented. For this reason, the control group was obtained from one hospital and included 19,372 patients from various hospital wards: allergology, angiology, surgery, phoniatics, gastroenterology, hematology, cardiology, neonatology, neurology, neurosurgery, ophthalmology, otolaryngology, gynecology, obstetrics, rheumatology, urology, anesthesiology, internal medicine, emergency medicine, orthopedics and outpatients. None of the patients in the control group had been diagnosed with FD, and there was no information about FD in their EHRs. Unfortunately, the control group could potentially include undiagnosed FD patients. Patients were selected from the Saventic Health database according to inclusion and exclusion criteria listed below. (i) Inclusion criteria: age 18–75; no other criteria were defined in order to imitate implementation of the decision-support scoring system in a hospital population. (ii) Exclusion criteria: palliative care, alcohol dependency, disseminated cancer, blood cancer, and chemotherapy.

2.2. FD Risk Factor Development

The goal of the presented approach was to develop a scoring system for FD diagnosis based on digitized data from hospital records and its practical implementation. Firstly, according to detailed literature screening [23,35,36] and consultations with physicians from FD knowledge centers, a set of FD-related signs available in the literature was brought to 13 clinical features, considered by physicians as most useful and typical for FD. Secondly, based on expert knowledge, a scoring system for 13 clinical features was created. The score of each feature ranged from zero to three. When creating a scoring system, the following aspects were taken into consideration: the number of traits and characteristics included in a particular feature; the significance of a feature in FD diagnosis and/or severity; and

the typical age of a feature's first manifestation. Therefore, when a feature was common in the general population at a specific age, its score was low. On the other hand, if a feature typical for the elderly was observed in younger patients, e.g., stroke, myocardial infarction, renal failure, eye fundus lesions, or hearing loss, the scoring weight increased the score value of a feature. The sum of the clinical feature scores of a patient constituted the FD risk factor. Next, some of the clinical features typical for FD may be caused by other diseases known to be relevant for a patient. In such cases, when another confirmed disease was a potential source of a clinical feature, then it was not scored. In order to achieve this, several feature exclusion criteria were considered (Figure 1). One exclusion criterion was concerned with renal failure: nephropathies, amyloidosis, sarcoidosis, Alport's syndrome, post-nephrectomy condition (single kidney, bilateral), metastatic urolithiasis, and congenital renal system defects. Another exclusion criterion was related with the clinical features of stroke: patent foramen ovale (PFO) and trauma. Finally, the 13 clinical features selected were further categorized according to the five most common FD signs: (1) cardiovascular symptoms; (2) kidney disease; (3) skin changes in selected areas; (4) neurological disorders; and (5) eye changes (Figure 1).

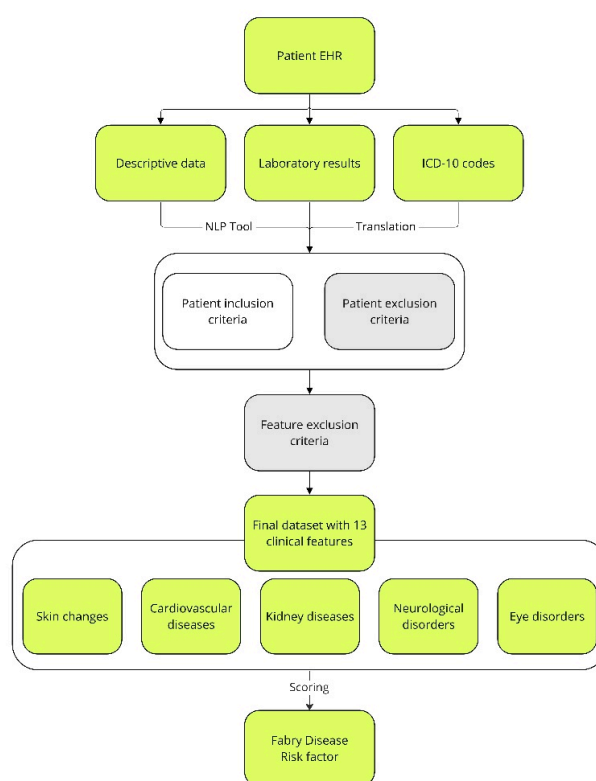


Figure 1. Overview of patients' assignment to the study and control groups, as well as the extraction of characteristics related to FD from text records and descriptions of ICD-10 codes. The fields that were used for FD risk factor assessment are indicated in lime. Fields eliminating patients from further analysis due to specific symptoms or exclusion criteria, which are described in the Materials and Methods, are indicated in gray. EHR—electronic health records; ICD-10—International Statistical Classification of Diseases and Related Health Problems; NLP—natural language processing.

2.3. Natural Language Processing

NLP was implemented based on spaCy's *pl_core_news_md* model [30]. The name *pl_core_news_md* consists of four elements: *pl* stands for the language on which the model was trained; *core* stands for general purpose, that is, tagging, parsing, lemmatization, and named entity recognition; *news* stands for the type of text data (blogs, news, com-

ments) on which the model was originally trained by the spaCy developers; and finally, *md* stands for medium package size. This size has a reduced word vector table with 20 thousand unique vectors, resulting in approximately 500 thousand words. In our study, NLP was used to evaluate text records in the EHRs to obtain detailed patient characteristics (Figure 2). Regular expressions were combined to describe a medical term (Figure 2, step 1). Then, extensive medical descriptions found in the EHRs were broken down into sentences (Figure 2, step 2). The next step was to correct errors and typos occurring in the sentences, as well as to perform lemmatization (Figure 2, step 3). The text prepared in this way was subjected to tokenization (Figure 2, step 4). Tokenization is the division of sentences into meaningful units, usually separated in the text by white space. Regular expressions created in step 1 were searched for in the token sequence (Figure 2, step 5). Tokens were evaluated for the presence of a given medical term or its negation (Figure 2, step 6). This stage was supervised and usually required several hundred analyzed descriptions before it was considered to have worked properly, and the term saved in the dictionary (Figure 2, step 7). Subsequently, the cosine similarity of the specified token to an expression that already exists in the dictionary was checked (Figure 2, step 8). Finally, the medical term was extracted and evaluated, which contributed to the feature score (Figure 2, step 9). In cases where a token was related to a medical term that already existed and was stored in the dictionary, steps 6 and 7 were omitted. To determine how well the extraction technique performed, precision-oriented tests were conducted. Symptom-wise, 100 medical descriptions were analyzed, in which the presence of a symptom was detected; then, the correctness of the findings was manually verified by previously trained annotators. The results varied among symptoms, but they all exceeded the 70% precision threshold, at least. The model did not identify abbreviations if they were not in the dictionary.

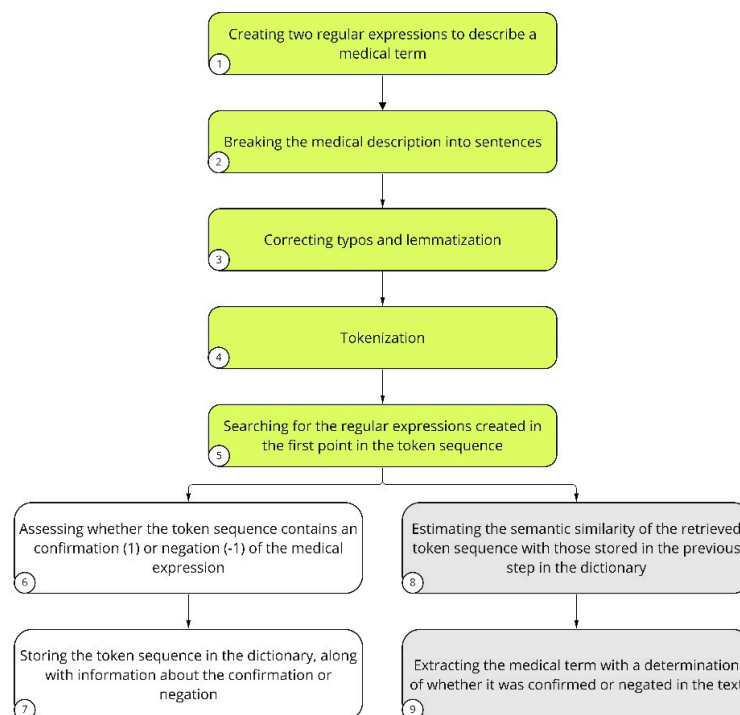


Figure 2. Description of NLP analysis performed on text records of the study and control patient groups. The steps of LP involving preparation of the patients’ EHRs to be assessed are indicated in green. The key NLP steps for introducing a new term to the dictionary are indicated in white. Medical terms that are already present in the dictionary are indicated in gray.

2.4. FD Risk Score Implementation

To assess the FD risk factor among the studied population of patients, their EHRs were extracted from hospital systems. In a first step, ICD-10 codes were turned into their full descriptions, which allowed the extraction of additional patient information. Next, with the use of NLP, detailed patient characteristics were obtained from the text records. Then, the NLP-determined traits, laboratory test results, and characteristics obtained from the ICD-10 codes were assigned to 13 clinical features. Finally, using the established scoring system, each patient was assigned an FD risk factor. For the obtained values of risk factor for the patients from the study and control groups, a distribution plot was created and a cut-off value was established. The purpose of the cut-off value was to effectively reduce the number of patients referred for screening.

Several cut-off values were tested in order to choose the one that maximized the detection specificity of FD patients. None of the patients from the study group achieved a risk factor lower than three. For this reason, 3 was the lowest considered value for the cut-off. The risk factor values were verified between the cut-off of 3 and 11 in order to determine the specificity for the detection of FD patients. Based on the chosen cut-off value, patients were divided into those having low and high risk factor scores. The EHRs of those with high risk factor scores were further evaluated by two physicians, who decided whether a patient should be reported to the coordinator at the hospital at which the patient was being treated. There, the attending physician, having received the report, made a decision as to whether to contact the patient for further verification, e.g., to refer for DBS assessment, as is required for final FD confirmation [37,38]. At the final step of the analytical pipeline, a confusion matrix was created based on the established cut-off point. It was used to evaluate the number of patients with high or low risk factor scores among the total number of patients in the study and control groups. The quality of the FD risk factor scoring system was evaluated as follows: (1) accuracy, Equation (1); (2) precision, Equation (2); (3) recall, Equation (3); (4) F1-score, Equation (4); where TN —non-FD patients correctly classified as non-FD; TP —FD patients correctly classified as FD patients; FP —non-FD patients misclassified as FD patients; FN —FD patients misclassified as non-FD.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 \text{ score} = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (4)$$

2.5. Statistical Analysis

Specific statistical tests were used to verify whether there was any evidence rejecting the null hypothesis, considering the attributes of the study and control groups. The null hypothesis was that there is no difference between the study and control groups. The Mann–Whitney test was used to compare continuous numerical variables, e.g., age, while the chi-squared test and Fisher’s exact test were used to compare categorical variables, e.g., sex proportions, incidence of cardiovascular diseases, skin changes, neurological disorders, kidney diseases, and eye disorders between the two analyzed groups. Statistics were calculated using the Sklearn [39] and Scipy [40] libraries, while the visualizations were created using Matplotlib [41] and Seaborn [42] libraries.

3. Results

This study presents a detailed NLP-based analysis of the EHRs of 19,385 patients. Their general characteristics and five most common FD signs are shown in Table 1. According to the obtained results, the prevalence of Fabry was higher among males than females (by 11.5%). The number of females and males was equal in the control group (within an accuracy of 1%). The statistical test failed to reveal any significant difference in terms of gender distribution between the study and control groups. Hence, there is not enough evidence to confirm a gender predisposition based on the observed differences among FD patients. In other words, patients are likely to have random chances rather than gender predisposition. On the other hand, the statistical test revealed a significant difference in mean age between the two groups, with FD patients being younger than patients in the control group. The results of the other tests showed that the distribution of traits differed significantly between both groups. As no eye disorders were observed in the study group, we excluded this sign from hypothesis testing. It was found that the most frequent sign in both the study and control groups was cardiovascular disease (Table 1) with the following clinical feature: hypertrophic.

Table 1. General characteristics and five most common FD signs, extracted with the use of NLP, presented for the patients from the study and control groups; Study group—patients with confirmed FD; Control group—patients presumably without FD. SD—standard deviation.

	Study Group (n = 13)	Control Group (n = 19,372)	p-Value
Sex (% female)	38.5%	50.5%	0.38
Mean age (SD)	45.2 (10.5)	55.5 (13.3)	p < 0.05
Cardiovascular diseases (%)	10 (76.9%)	6641 (34.3%)	p < 0.05
Skin changes (%)	8 (61.5%)	149 (0.8%)	p < 0.05
Neurological disorders (%)	7 (53.8%)	5034 (26.0%)	p < 0.05
Kidney diseases (%)	4 (30.8%)	2030 (10.5%)	p < 0.05
Eye disorder (%)	0 (0%)	298 (1.5%)	-

The frequency of the 13 clinical features selected was scored to obtain the risk factor, and this varied between groups (Figure 3). Of all of the clinical features taken into account as part of the FD risk factor, the most frequent in the study group was myocardial infarction (Figure 3), while in the control group, it was hypertrophic cardiomyopathy (Figure 3).

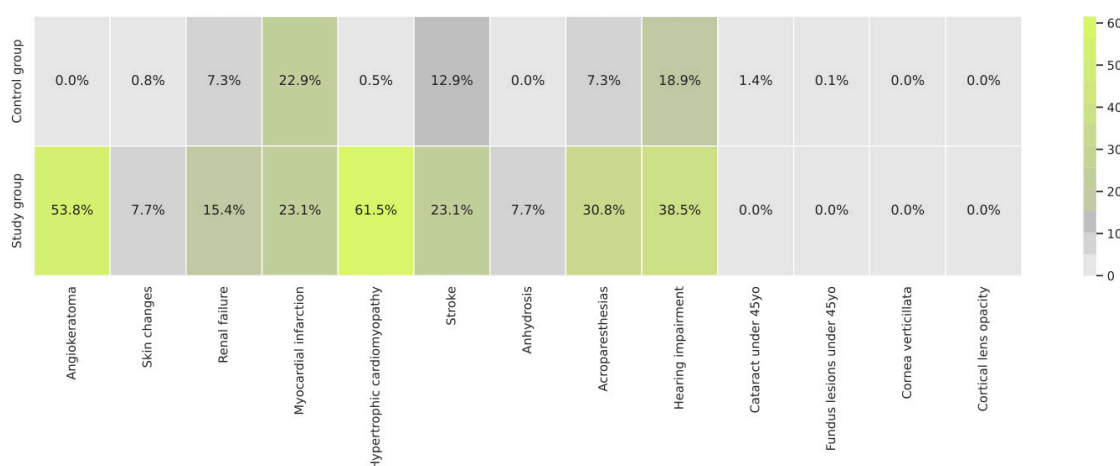


Figure 3. Frequency of the clinical features extracted from EHRs with the use of NLP in the study and control groups.

The risk factor was individually computed for each patient from the study and control groups based on the NLP analysis performed on their EHRs (Figure 4). Using the obtained results, a distribution of the number of patients with a given risk factor was plotted. The patients with confirmed FD in the study group were aligned following the ranking, showing high risk score values (Figure 4).

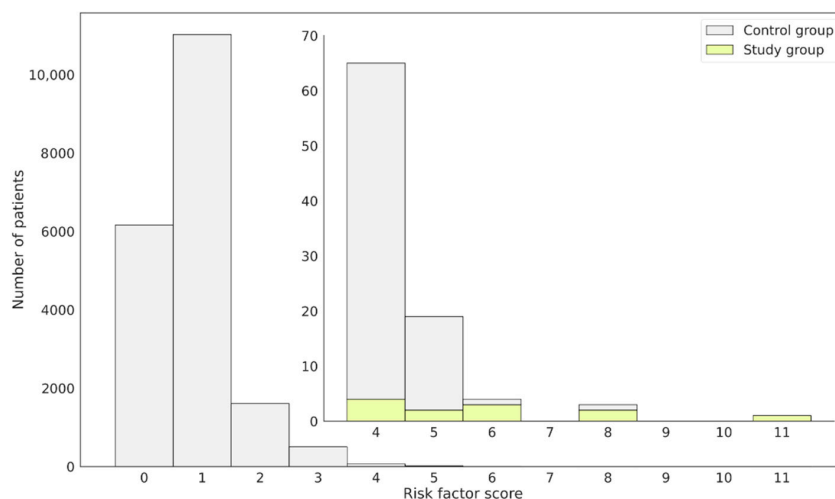


Figure 4. Distribution of patients assigned to the control or study group regarding the obtained FD risk factor score. Confirmed FD patients, constituting the study group, obtained high-risk score values and were placed on the right-hand side of the plot. The external bars represent the whole patient population, while the internal focus on those having the risk factor ≥ 4 . Bars in yellow represent.

Patients with Diagnosed FD

Based on the distribution of the number of patients with specific risk factor values, several cut-offs were examined. For the assumed cut-off of 3, the specificity obtained was lower than when the cut-off was set to 4. At the same time, when assuming a cut-off equal to 5 or higher, the specificity increased, but the sensitivity dropped drastically. For this reason, a cut-off value equal to 4 was chosen. This effectively reduced the number of patients referred for screening. An adjustment was made for the risk factor of patients with confirmed FD, so that they would be included among those referred for screening. Moreover, the determined cut-off value accurately separated patients who should be referred for diagnostic tests and those whose symptoms were not alarming and/or had no signs of FD (Figure 4). In total, there were 92 patients who obtained a high risk factor ≥ 4 . The risk factor distributions in the control and study groups are presented in Figure 5.

Out of 92 patients with risk factors above the cut-off point, 80 were initially assigned to the control group, and their average risk factor was 4.3 (SD 0.62). The EHRs of these patients were verified by two physicians experienced in the diagnosis of FD, who came to the conclusion that of the 80 patients, 15 should be reported to the coordinators of the hospitals in which they were being treated, so that their attending physicians could decide whether to refer them for DBS assay for further diagnosis. At this point, it is known that one patient originally assigned to the control group, who received a high risk factor equal to 8 (Figure 4), was diagnosed with FD after a DBS assay was performed. Other patients received a negative test result. It is worth noting that there was one patient in the study group who was diagnosed with FD with a risk factor of 3, which is below the chosen cut-off point. The EHR of this patient was manually verified, and it was found that some symptoms, like skin changes, were described as: “Lesions of the type of multiple cavernous angiomas on the skin. Long-standing problem. Undiagnosed”, which, due to the general indications, was rated low in the scoring system, but if assigned more directly as “angiokeratoma” would have resulted in a higher risk factor assignment.

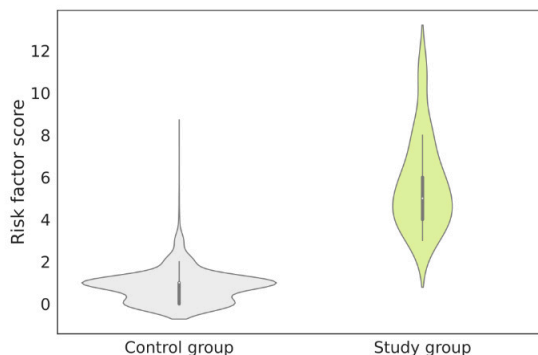


Figure 5. Risk factor distributions in the control group and the study group. The mean risk factors in the study group and the control group were 5.7 (SD 2.13) and 0.8 (SD 0.72), respectively. SD—standard deviation.

As mentioned, there were 92 patients with the risk factor ≥ 4 . Among them, 80 patients were classified as FP and 12 were TP, which gave the ratio 1:7, while in the general population it is known to be 1:40,000 [1]. The obtained 1:7 ratio is 5714 times higher than the ratio represented within the general population. This represents significant knowledge gain. Next, a confusion matrix on all patients included in this study was prepared (Figure 6) to summarize the classification performance of the presented risk factor scoring for the cut-off set to 4. That evaluation helped to select patients who disagreed with the original class. The only FN patient known to have FD was confirmed to have a very general description in the EHR (as already mentioned). The returned confusion matrix (Figure 6) indicated an accuracy of 0.996, with sensitivity = 92.3% and specificity = 99.59%. However, due to the high class imbalance, precision = 0.1304, area under the precision–recall curve = 0.537 and F1-score = 0.2286 were found to be much more informative metrics. The obtained recall value supported the project’s assumptions. The low value of precision was anticipated, as from the beginning, it was assumed that some FD patients could be present in the control group. The high AUC value of 0.998 confirmed that the applied risk factor scoring and assigned cut-off point enabled accurate distinction between patient groups with a high discrimination power.

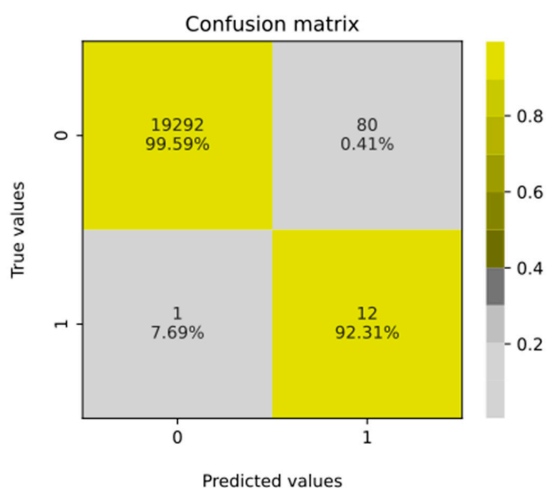


Figure 6. Confusion matrix computed for all patients including their FD risk factor values and cut-off value equal to 4. When interpreting the given values, it is important to remember that the issue concerns the prevalence of a rare disease (see Discussion).

4. Discussion

Despite the low prevalence of rare diseases, their proper and early diagnosis, although challenging, is critical for affected patients. Therefore, a machine learning approach can accelerate their appropriate determination, e.g., in the case of FD [23,43,44], particularly when facilitated by NLP, as described in this paper. In this case, EHR analysis can highlight seemingly imperceptible facts or symptoms that typically elude physicians, preventing proper FD detection, especially in the case of physicians who have never seen the symptoms of this disease, e.g., cherry-like skin changes (angiokeratoma) coupled with pain in the extremities and recurrent fever. Furthermore, since FD symptoms develop over many years, long-term monitoring of medical history is essential [45]. Therefore, the long-term incorporation of AI algorithms may be a valuable tool for supporting physicians in their clinical practice. These algorithms will increase the chance of detecting typical symptoms, characteristic combinations thereof, and their changes over time, leading to correct diagnosis. Referring people with high risk factor, who might potentially suffer from FD, for diagnostic testing may be more effective and less expensive than screening the whole population. Moreover, AI algorithms that support physicians may also increase the awareness of rare diseases in the medical community. As a next step, they might also improve therapeutic approaches and prognostic decision-making processes in affected individuals. We presented a novel NLP-based approach that can support physicians in the diagnosis of FD based on the medical information included in patients' EHRs, covering symptoms, laboratory results, and ICD-10 codes. The patients in the study group were 23–59 years old; thus, we decided that only adults would be included in the control group. In addition, the upper age limit in the control group was 75 years, in line with [46], where it was reported that more than 99% of 2044 FD patients had been diagnosed with FD before that age. As a result of this study, a narrow group of patients with a high risk factor of FD was identified. The likelihood of FD diagnosis in patients whose risk factor score exceeded the cut-off value was several thousand times higher than in the general population. The presented approach showed very high accuracy. However, accuracy might be a misleading metric in the case of imbalanced datasets where the searched trait is characterized by a very low incidence, as is the case with rare diseases. The disparity between the numbers of patients in the study and control groups was not related to data preparation, but to the fact that FD is a rare disease. For this reason, the AUC, precision, and F1-score served as performance metrics for the presented approach. In order to evaluate the effectiveness of the presented FD risk scoring system, a simple comparison was performed between the AUC values of the presented approach (AUC 0.998) and the approach designed by Jeffries et al. [23] (AUC 0.82). In the present study, the number of patients in the control group who had risk factor ≥ 4 was relatively small. Patients were assessed for a specific clinical presentation, and whether there were FN patients in the control group who did not undergo DBS assay remains uncertain. Therefore, it is difficult to judge what the true AUC was and whether it was really higher than the 0.82 presented by Jeffries et al. [23]. Their study [23] had significantly more participants, with close to 5000 FD patients in the study group and 1,000,000 patients in the control group, giving a ratio between study/control equal to 0.005. In our study, this ratio was 0.00067. Due to this fact, it is difficult to clearly determine which method is more effective. The presented approach is not a classic example of a classifier. Initially, the study group consisted of only 13 patients diagnosed with FD. Modeling such a small sample is difficult because of the high risk of overfitting due to the limited model verification capabilities. Test–train partitioning is affected by the sampling mechanism, and model learning and evaluation methods such as bootstrapping or leave-one-out cross-validation are computationally expensive. Models resulting from small-sample modeling are simple in structure, and factor-based models provide an alternative characterized by full controllability and explainability. Therefore, a risk-factor-based approach was used to select patients who should be referred for specialized diagnostic testing. The presented approach was developed to screen patients at the population scale. This would facilitate and accelerate the work of physicians. Moreover, the use of the presented approach could

potentially reduce diagnostic costs, as it will indicate the need for diagnostic testing only for patients whose risk factor score exceeds the cut-off value.

The implementation of the presented approach contributed to the diagnosis of one patient with FD, as confirmed with DBS assay. This patient was 45, with a history of unexplained cardiomyopathy, vertigo, asymptomatic ischemic changes in the brain MRI, and renal transplant in a first-degree relative under 45, resulting in a risk factor of 8. At the time of the implementation of the risk factor-based approach, this patient was clinically suspected of amyloidosis, and FD was not considered as a potential reason at any point in the patient's EHRs. Thus, our method led to the proper diagnosis, which is critical for the affected patient. Family screening also confirmed a diagnosis of FD in the patient's brother. The approach presented in this paper makes it possible to narrow the pool of patients with a high risk factor who should be referred for DBS assay. Even imperfect indications can be perceived as beneficial due to the raising of physicians' suspicions, followed by patients' referring for further evaluation and proper diagnostic tests, therefore resulting in earlier diagnosis. However, it should be kept in mind that this method of risk factor assessment still has a chance of missing individual patients whose EHRs analysis incorrectly underestimates their risk factor. However, in such cases, further data follow-up may be decisive. The use of the approach presented in this paper may prove particularly useful due to the fact that it is able to draw attention to seemingly irrelevant facts and can combine multiple features (e.g., ICD-10 codes) to improve predictive and diagnostic capabilities. In recent years, we have observed a breakthrough in NLP competition, large pretrained transformers like BERT [47] and GPTs [48] have become state-of-art and general-purpose language models due to their capacity to contextualize word representations. This means that models like BERT can understand the meaning of words based on their context in a sentence, allowing it to capture more nuanced and complex relationships between words. It has been demonstrated that BERT-based models can be used effectively in medical tasks [49]. Unsupervised training generates embeddings used as inputs for downstream supervised tasks such as disease classification. They can significantly improve performance, particularly on smaller datasets where labeled data are limited. The vast majority of implements have considered only the English language; however, our research is based on content in Polish. To overcome this problem, procedures for transferring knowledge from multilingual to monolingual BERT-based models were developed, thanks to which HerBERT (Polish BERT) was developed [50]. Transformers constitute the state of the art, and can be considered to be a direction worth incorporating into our study to enhance or replace the NLP search approach we developed.

5. Conclusions

This work aimed to create a risk factor scoring system to support physicians in FD diagnosis using real-life data. The authors used an original approach to the NLP tool evaluating patients' EHRs and assessing risk factor among patients who may suffer from FD. The proposed method increases the effectiveness of diagnostics, further improving patients' quality of life and prognosis. To the best of the authors' knowledge, this is the first case of FD diagnosis with the help of NLP application. Despite the promising results of this experiment, there are several limitations that need to be considered. First of all, the sample size of the FD patients was relatively small, and EHRs are incomplete, which may limit the generalizability of the conclusions drawn. Therefore, further development and testing of the risk factor with a larger and more diverse patient population are required to precisely assess usability in a clinical setting. Secondly, the control group could potentially include undiagnosed FD patients. However, for this reason, we present our solution as a screening tool at the population level, and not as a classifier in itself. Thirdly, the NLP algorithm implemented in this study requires further development to improve its accuracy. While the algorithm has demonstrated its utility, there is still room for improvement in terms of its ability to generalize and analyze context. Since the confirmation of the studies are prospective tests, an appropriate amount of time is needed to collect patients

for the DBS examinations. These limitations highlight the need for continued research and development.

Author Contributions: Conceptualization: A.A.M., K.L. and G.W.B.; Data curation: K.L. and J.S.; EHRs verification: K.L. and J.S.; Formal analysis: A.A.M. and A.S.; Investigation: A.A.M. and A.S.; Methodology: A.A.M., K.L., J.S., S.M.K. and M.J.D.; Software: A.A.M. and A.S.; Supervision: K.L. and M.J.D.; Validation: A.A.M. and A.S.; Visualization: A.A.M.; Writing—Original draft and editing: A.A.M., S.M.K. and M.J.D.; Writing—Review: M.D., K.M.-S., M.N., S.B.-S., G.W.B., A.S., K.L. and M.J.D. All authors have read and agreed to the published version of the manuscript.

Funding: The source of funding did not affect the data collection, analysis, or the results described in this manuscript. The authors confirm that they had full access to all of the data in this study and take complete responsibility for the integrity of the data and the accuracy of the data analysis.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the or Ethics Committee of Collegium Medicum in Bydgoszcz Nicolaus Copernicus University in Torun, Poland (protocol code KB 487/2022 on 27 September 2022).

Informed Consent Statement: The informed consent of the patients was not needed to perform analysis based on medical data from health care centers.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to personal data and privacy protection.

Acknowledgments: The authors express special thanks to Marta Osińska for verifying the linguistic correctness of an earlier version of this manuscript, Jan Biziel University Hospital No. 2 in Bydgoszcz, Poland, who contributed to the verification of patients at risk of Fabry disease, Polish Association of Families with Fabry Disease and Fabry disease patients who decided to share their EHRs.

Conflicts of Interest: The authors whose names are listed below report the following details of affiliation with a financial interest in the subject matter or materials discussed in this manuscript: Adrian A. Michalski—Data Scientist Lead in Saventic Health, Polna 66/12 Street, 87-100 Torun, Poland; Karol Lis—Medical Director in Saventic Health, Polna 66/12 Street, 87-100 Torun, Poland; Joanna Stankiewicz—Medical Project Lead in Saventic Health, Polna 66/12 Street, 87-100 Torun, Poland; Sylwester M. Kloska—Data Analyst in Saventic Health, Polna 66/12 Street, 87-100 Torun, Poland; Arkadiusz Sycz—Machine Learning Engineer in Saventic Health, Polna 66/12 Street, 87-100 Torun, Poland; Marek Dudziński—Clinical Advisor in Saventic Health, Polna 66/12 Street, 87-100 Torun, Poland; Katarzyna Muras-Szwedziak—Clinical Advisor at Saventic Foundation, Polna 66 Street, 87-100 Torun, Poland; Michał Nowicki—Clinical Advisor at Saventic Foundation, Polna 66 Street, 87-100 Torun, Poland; Stanisława Bazan-Socha—Clinical Advisor at Saventic Foundation, Polna 66 Street, 87-100 Torun, Poland; Michał J. Dabrowski—Publication & Grants Lead in Saventic Health, Polna 66/12 Street, 87-100 Torun, Poland; Grzegorz W. Basak—CMO, Founder of Saventic Health, Polna 66/12 Street, 87-100 Torun, Poland.

References

1. Mehta, A.; Beck, M.; Eyskens, F.; Feliciani, C.; Kantola, I.; Ramaswami, U.; Rolfs, A.; Rivera, A.; Waldek, S.; Germain, D. Fabry disease: A review of current management strategies. *QJM Int. J. Med.* **2010**, *103*, 641–659. [[CrossRef](#)] [[PubMed](#)]
2. Garman, S.C.; Garboczi, D.N. The molecular defect leading to Fabry disease: Structure of human α -galactosidase. *J. Mol. Biol.* **2004**, *337*, 319–335. [[CrossRef](#)] [[PubMed](#)]
3. Aerts, J.M.; Groener, J.E.; Kuiper, S.; Donker-Koopman, W.E.; Strijland, A.; Ottenhoff, R.; van Roomen, C.; Mirzaian, M.; Wijburg, F.A.; Linthorst, G.E.; et al. Elevated globotriaosylsphingosine is a hallmark of Fabry disease. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 2812–2817. [[CrossRef](#)] [[PubMed](#)]
4. Hwu, W.L.; Chien, Y.H.; Lee, N.C.; Chiang, S.C.; Dobrovolny, R.; Huang, A.C.; Yeh, H.Y.; Chao, M.C.; Lin, S.J.; Kitagawa, T.; et al. Newborn screening for Fabry disease in Taiwan reveals a high incidence of the later-onset GLA mutation c. 936+ 919G> A (IVS4+ 919G> A). *Hum. Mutat.* **2009**, *30*, 1397–1405. [[CrossRef](#)]
5. MacDermot, K.; Holmes, A.; Miners, A. Anderson-Fabry disease: Clinical manifestations and impact of disease in a cohort of 98 hemizygous males. *J. Med. Genet.* **2001**, *38*, 750–760. [[CrossRef](#)]
6. Mechtler, T.P.; Stary, S.; Metz, T.F.; De Jesús, V.R.; Greber-Platzer, S.; Pollak, A.; Herkner, K.R.; Streubel, B.; Kasper, D.C. Neonatal screening for lysosomal storage disorders: Feasibility and incidence from a nationwide study in Austria. *Lancet* **2012**, *379*, 335–341. [[CrossRef](#)]

7. Spada, M.; Pagliardini, S.; Yasuda, M.; Tukul, T.; Thiagarajan, G.; Sakuraba, H.; Ponzzone, A.; Desnick, R.J. High incidence of later-onset Fabry disease revealed by newborn screening. *Am. J. Hum. Genet.* **2006**, *79*, 31–40. [\[CrossRef\]](#)
8. Di Risi, T.; Vinciguerra, R.; Cuomo, M.; Della Monica, R.; Riccio, E.; Cocozza, S.; Imbriaco, M.; Duro, G.; Pisani, A.; Chiariotti, L. DNA methylation impact on Fabry disease. *Clin. Epigenetics* **2021**, *13*, 24. [\[CrossRef\]](#)
9. Echevarria, L.; Benistan, K.; Toussaint, A.; Dubourg, O.; Hagege, A.; Eladari, D.; Jabbour, F.; Beldjord, C.; De Mazancourt, P.; Germain, D. X-chromosome inactivation in female patients with Fabry disease. *Clin. Genet.* **2016**, *89*, 44–54. [\[CrossRef\]](#)
10. Germain, D.P.; Oliveira, J.P.; Bichet, D.G.; Yoo, H.W.; Hopkin, R.J.; Lemay, R.; Politei, J.; Wanner, C.; Wilcox, W.R.; Warnock, D.G. Use of a rare disease registry for establishing phenotypic classification of previously unassigned GLA variants: A consensus classification system by a multispecialty Fabry disease genotype–phenotype workgroup. *J. Med. Genet.* **2020**, *57*, 542–551. [\[CrossRef\]](#)
11. Politei, J.M.; Bouhassira, D.; Germain, D.P.; Goizet, C.; Guerrero-Sola, A.; Hilz, M.J.; Hutton, E.J.; Karaa, A.; Liguori, R.; Üçeyler, N.; et al. Pain in Fabry disease: Practical recommendations for diagnosis and treatment. *CNS Neurosci. Ther.* **2016**, *22*, 568–576. [\[CrossRef\]](#) [\[PubMed\]](#)
12. MacDermot, K.; Holmes, A.; Miners, A. Natural history of Fabry disease in affected males and obligate carrier females. *J. Inherit. Metab. Dis.* **2001**, *24*, 13. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Mehta, A.; Ricci, R.; Widmer, U.; Dehout, F.; Garcia de Lorenzo, A.; Kampmann, C.; Linhart, A.; Sunder-Plassmann, G.; Ries, M.; Beck, M. Fabry disease defined: Baseline clinical manifestations of 366 patients in the Fabry Outcome Survey. *Eur. J. Clin. Investig.* **2004**, *34*, 236–242. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Reisin, R.; Perrin, A.; García-Pavía, P. Time delays in the diagnosis and treatment of Fabry disease. *Int. J. Clin. Pract.* **2017**, *71*, e12914. [\[CrossRef\]](#)
15. Rosa Neto, N.S.; de Barros Bento, J.C.; Pereira, R.M.R. Higher rate of rheumatic manifestations and delay in diagnosis in Brazilian Fabry disease patients. *Adv. Rheumatol.* **2020**, *60*, 7. [\[CrossRef\]](#)
16. Thomas, A.S.; Mehta, A.B. Difficulties and barriers in diagnosing Fabry disease: What can be learnt from the literature? *Expert Opin. Med. Diagn.* **2013**, *7*, 589–599. [\[CrossRef\]](#)
17. Nestler-Parr, S.; Korchagina, D.; Toumi, M.; Pashos, C.L.; Blanchette, C.; Molsen, E.; Morel, T.; Simoens, S.; Kaló, Z.; Gattermann, R.; et al. Challenges in research and health technology assessment of rare disease technologies: Report of the ISPOR rare disease special interest group. *Value Health* **2018**, *21*, 493–500. [\[CrossRef\]](#)
18. Nowicki, M.; Bazan-Socha, S.; Blazejewska-Hyzorek, B.; Gellert, R.; Imiela, J.; Ka'zmierzak, J.; Kłopotowski, M.; Oko-Sarnowska, Z.; Pawlaczek, K.; Ponikowski, P.; et al. Enzyme replacement therapy in Fabry disease in Poland: Position statement. *Pol. Arch. Intern. Med.* **2020**, *130*, 91–97. [\[CrossRef\]](#)
19. Germain, D.P.; Charrow, J.; Desnick, R.J.; Guffon, N.; Kempf, J.; Lachmann, R.H.; Lemay, R.; Linthorst, G.E.; Packman, S.; Scott, C.R.; et al. Ten-year outcome of enzyme replacement therapy with agalsidase beta in patients with Fabry disease. *J. Med. Genet.* **2015**, *52*, 353–358. [\[CrossRef\]](#)
20. Rombach, S.M.; Smid, B.E.; Bouwman, M.G.; Linthorst, G.E.; Dijkgraaf, M.G.; Hollak, C.E. Long term enzyme replacement therapy for Fabry disease: Effectiveness on kidney, heart and brain. *Orphanet J. Rare Dis.* **2013**, *8*, 47. [\[CrossRef\]](#)
21. Weidemann, F.; Sanchez-Niño, M.D.; Politei, J.; Oliveira, J.P.; Wanner, C.; Warnock, D.G.; Ortiz, A. Fibrosis: A key feature of Fabry disease with potential therapeutic implications. *Orphanet J. Rare Dis.* **2013**, *8*, 116. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Hughes, D.A.; Malmenäs, M.; Deegan, P.B.; Elliott, P.M.; Ginsberg, L.; Hajioff, D.; Ioannidis, A.S.; Orteu, C.H.; Ramaswami, U.; West, M.; et al. Fabry International Prognostic Index: A predictive severity score for Anderson-Fabry disease. *J. Med. Genet.* **2012**, *49*, 212–220. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Jefferies, J.L.; Spencer, A.K.; Lau, H.A.; Nelson, M.W.; Giuliano, J.D.; Zabinski, J.W.; Boussios, C.; Curhan, G.; Gliklich, R.E.; Warnock, D.G. A new approach to identifying patients with elevated risk for Fabry disease using a machine learning algorithm. *Orphanet J. Rare Dis.* **2021**, *16*, 518. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Wang, J.; Abu-el Rub, N.; Gray, J.; Pham, H.A.; Zhou, Y.; Manion, F.J.; Liu, M.; Song, X.; Xu, H.; Rouhizadeh, M.; et al. COVID-19 SignSym: A fast adaptation of a general clinical NLP tool to identify and normalize COVID-19 signs and symptoms to OMOP common data model. *J. Am. Med. Inform. Assoc.* **2021**, *28*, 1275–1283. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Neuraz, A.; Lerner, I.; Digan, W.; Paris, N.; Tsopra, R.; Rogier, A.; Baudoin, D.; Cohen, K.B.; Burgun, A.; Garcelon, N.; et al. Natural language processing for rapid response to emergent diseases: Case study of calcium channel blockers and hypertension in the COVID-19 pandemic. *J. Med. Internet Res.* **2020**, *22*, e20773. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Lybarger, K.; Ostendorf, M.; Thompson, M.; Yetisgen, M. Extracting COVID-19 diagnoses and symptoms from clinical text: A new annotated corpus and neural event extraction framework. *J. Biomed. Inform.* **2021**, *117*, 103761. [\[CrossRef\]](#)
27. Chapman, A.; Peterson, K.; Turano, A.; Box, T.; Wallace, K.; Jones, M. A Natural Language Processing System for National COVID-19 Surveillance in the US Department of Veterans Affairs. In Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020, Online, 9 July 2020; Association for Computational Linguistics: Toronto, ON, Canada, 2020.
28. Schöning, V.; Liakoni, E.; Drewe, J.; Hammann, F. Automatic identification of risk factors for SARS-CoV-2 positivity and severe clinical outcomes of COVID-19 using Data Mining and Natural Language Processing. *medRxiv* **2021**. [\[CrossRef\]](#)
29. Oyebo, O.; Ndulue, C.; Adib, A.; Mulchandani, D.; Suruliraj, B.; Orji, F.A.; Chambers, C.T.; Meier, S.; Orji, R. Health, psychosocial, and social issues emanating from the COVID-19 pandemic based on social media comments: Text mining and thematic analysis approach. *JMIR Med. Inform.* **2021**, *9*, e22734. [\[CrossRef\]](#)

30. Honnibal, M.; Montani, I.; Van Landeghem, S.; Boyd, A. spaCy: Industrial-Strength Natural Language Processing in Python. Available online: <https://github.com/explosion/spaCy> (accessed on 6 January 2022).
31. Kluyver, T.; Ragan-Kelley, B.; Pérez, F.; Granger, B.; Bussonnier, M.; Frederic, J.; Kelley, K.; Hamrick, J.; Grout, J.; Corlay, S.; et al. Jupyter Notebooks—A publishing format for reproducible computational workflows. In *Proceedings of the Positioning and Power in Academic Publishing: Players, Agents and Agendas, Göttingen, Germany, 7–9 June 2016*; Loizides, F., Schmidt, B., Eds.; IOS Press: Amsterdam, The Netherlands, 2016; pp. 87–90.
32. Van Rossum, G.; Drake, F.L. *Python 3 Reference Manual*; CreateSpace: Scotts Valley, CA, USA, 2009.
33. Harris, C.R.; Millman, K.J.; van der Walt, S.J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N.J.; et al. Array programming with NumPy. *Nature* **2020**, *585*, 357–362. [[CrossRef](#)]
34. Pandas Development Team. pandas-dev/pandas: Pandas. Available online: <https://github.com/pandas-dev/pandas> (accessed on 6 January 2022).
35. Ortiz, A.; Germain, D.P.; Desnick, R.J.; Politei, J.; Mauer, M.; Burlina, A.; Eng, C.; Hopkin, R.J.; Laney, D.; Linhart, A.; et al. Fabry disease revisited: Management and treatment recommendations for adult patients. *Mol. Genet. Metab.* **2018**, *123*, 416–427. [[CrossRef](#)]
36. Tuttolomondo, A.; Pecoraro, R.; Simonetta, I.; Miceli, S.; Pinto, A.; Licata, G. Anderson-Fabry disease: A multiorgan disease. *Curr. Pharm. Des.* **2013**, *19*, 5974–5996. [[CrossRef](#)] [[PubMed](#)]
37. Delarosa-Rodríguez, R.; Santotoribio, J.D.; Paula, H.A.; González-Meneses, A.; García-Morillo, S.; Jiménez-Arriscado, P.; Guerrero, J.M.; Macher, H.C. Accuracy diagnosis improvement of Fabry disease from dried blood spots: Enzyme activity, lyso-Gb3 accumulation and GLA gene sequencing. *Clin. Genet.* **2021**, *99*, 761–771. [[CrossRef](#)] [[PubMed](#)]
38. Scalia, S. DBS assay in the diagnosis of Fabry disease. *Giornale Di Tecniche Nefrologiche E Dialitiche* **2017**, *29*, S5–S6. [[CrossRef](#)]
39. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
40. Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17*, 261–272. [[CrossRef](#)]
41. Hunter, J.D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95. [[CrossRef](#)]
42. Waskom, M.L. seaborn: Statistical data visualization. *J. Open Source Softw.* **2021**, *6*, 3021. [[CrossRef](#)]
43. Faviez, C.; Chen, X.; Garcelon, N.; Neuraz, A.; Knebelmann, B.; Salomon, R.; Lyonnet, S.; Saunier, S.; Burgun, A. Diagnosis support systems for rare diseases: A scoping review. *Orphanet J. Rare Dis.* **2020**, *15*, 94. [[CrossRef](#)]
44. Schaefer, J.; Lehne, M.; Schepers, J.; Prasser, F.; Thun, S. The use of machine learning in rare diseases: A scoping review. *Orphanet J. Rare Dis.* **2020**, *15*, 145. [[CrossRef](#)]
45. Lee, S.N.; Kim, G.H.; Yoo, K.D. A Case of a 50-Year-Old Woman with Typical Fabry Disease Who Showed Serial Electrocardiographic and Echocardiographic Changes over a 17-Year Period. *Case Rep. Cardiol.* **2019**, *2019*, 9385361.
46. Lidove, O.; Barbey, F.; Niu, D.M.; Brand, E.; Nicholls, K.; Bizjajeva, S.; Hughes, D.A. Fabry in the older patient: Clinical consequences and possibilities for treatment. *Mol. Genet. Metab.* **2016**, *118*, 319–325. [[CrossRef](#)] [[PubMed](#)]
47. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
48. Radford, A.; Narasimhan, K.; Salimans, T.; Sutskever, I. Improving Language Understanding by Generative Pre-Training. 2018. Available online: <https://openai.com/research/language-unsupervised> (accessed on 6 January 2022).
49. Rasmy, L.; Xiang, Y.; Xie, Z.; Tao, C.; Zhi, D. Med-BERT: Pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. *NPJ Digit. Med.* **2021**, *4*, 86. [[CrossRef](#)] [[PubMed](#)]
50. Mroczkowski, R.; Rybak, P.; Wróblewska, A.; Gawlik, I. HerBERT: Efficiently pretrained transformer-based language model for Polish. *arXiv* **2021**, arXiv:2105.01735.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Załącznik III

Oświadczenia współautorów publikacji
oraz zgoda Komisji Bioetycznej

Toruń, dnia 06.05.2024

dr inż. Konrad Duraj

Katedra Biosensorów i Przetwarzania Sygnałów Biomedycznych
Wydział Inżynierii Biomedycznej, Politechnika Śląska w Gliwicach

**Rada Dyscypliny Nauk Medycznych
Uniwersytetu Mikołaja Kopernika w
Toruniu**

Oświadczenie o współautorstwie

Niniejszym oświadczam, że w pracy Michalski A, Duraj K, Kupcewicz B. *Leukocyte deep learning classification assessment using Shapley additive explanations algorithm*. Int J Lab Hematol. 2023; 1-6; <https://doi.org/10.1111/ijlh.14031> mój udział polegał na implementacji algorytmów przetwarzania obrazów oraz rewizji manuskryptu. Mój udział w powstaniu pracy wynosi 20%.


(podpis)

Informacja o przetwarzaniu danych osobowych

Na podstawie Rozporządzenia Parlamentu Europejskiego i Rady (UE) 2016/679 z dnia 27 kwietnia 2016 r. w sprawie ochrony osób fizycznych w związku z przetwarzaniem danych osobowych i w sprawie swobodnego przepływu takich danych oraz uchylenia dyrektywy 95/46/WE, zwanego dalej „RODO”, informujemy, że:

1. Administratorem Pana/Pani danych osobowych będzie Uniwersytet Mikołaja Kopernika w Toruniu z siedzibą przy ul. Gagarina 11, 87-100 Toruń (dalej: Uczelnia, ADO).
2. Pana/Pani dane osobowe w związku z postępowaniem o nadanie stopnia naukowego będą przetwarzane:
 - a. na podstawie art. 6 ust. 1 lit. c) RODO – obowiązek prawny wynikający z przepisów ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (Dz. U. z 2020, poz. 85);
 - b. na podstawie art. 6 ust. 1 lit. f) RODO – prawnie uzasadniony interes ADO:
 - do ustalenia, obrony lub dochodzenia roszczeń – przez okres do przedawnienia roszczeń, lub przez okres prowadzenia postępowania przez właściwe organy lub sądy w przypadku dochodzenia roszczeń,
 - na potrzeby wewnętrzne tworzenia zestawień, analiz i statystyk – przez okres obowiązywania umowy,
 - na potrzeby prowadzenia ewidencji korespondencji przychodzącej i wychodzącej – wieceście,
 - na potrzeby marketingu produktów i usług ADO.
3. Z zastrzeżeniem przepisów powszechnie obowiązującego prawa przysługują Panu/Pani prawa, które zrealizujemy na wniosek o:
 - a. Żądanie dostępu do danych osobowych oraz prawo ich sprostowania,
 - b. Żądanie usunięcia lub ograniczenia przetwarzania.
4. Przysługuje Panu/Pani również prawo wniesienia sprzeciwu na przetwarzanie danych osobowych.
5. Podanie przez Pana/Panią danych osobowych jest niezbędne do wykonania celu wymienionego w pkt 2 lit. a i b., a brak ich podania uniemożliwi otwarcie i przeprowadzenie postępowania.
6. Przysługuje Panu/Pani prawo wniesienia skargi do Prezesa Urzędu Ochrony Danych Osobowych.
7. Pana/Pani dane osobowe mogą być udostępnione recenzentom lub organom administracji publicznej, sądom, komornikom w zakresie sytuacji przewidzianych w przepisach prawa.
8. Na dzień zbierania Pana/Pani danych osobowych nie planujemy przekazywać ich poza EOG (obejmujący Unię Europejską, Norwegię, Lichtenstein i Islandię), nie wykluczając tego w przyszłości, o czym zostanie Pan/Pani poinformowania ze stosownym wyprzedzeniem.
9. W stosunku do Pana/Pani nie będą prowadzone działania polegające na podejmowaniu decyzji w sposób zautomatyzowany, nie będą one również podlegały zautomatyzowanemu profilowaniu.
10. Jeżeli chce Pan/Pani skontaktować się z Uczelnią w sprawach związanych z przetwarzaniem danych osobowych, w szczególności w związku z wniesieniem wniosku o realizację przysługujących praw prosimy o kontakt pod adresem e-mail: iod@umk.pl lub adresem korespondencyjnym: UMK w Toruniu, ul. Gagarina 11, 87-100 Toruń, z dopiskiem „IOD”, dostępny jest również kontakt telefoniczny: 56 611 27 42.

Toruń, dnia 06.05.2024

dr hab. Bogumila Kupcewicz, prof. UMK

Katedra Chemii Nieorganicznej i Analitycznej
Wydział Farmaceutyczny CM, Uniwersytet Mikołaja Kopernika

**Rada Dyscypliny Nauk Medycznych
Uniwersytetu Mikołaja Kopernika w
Toruniu**

Oświadczenie o współautorstwie

Niniejszym oświadczam, że w pracy Michalski A, Duraj K, Kupcewicz B. *Leukocyte deep learning classification assessment using Shapley additive explanations algorithm*. Int J Lab Hematol. 2023; 1-6; <https://doi.org/10.1111/ijlh.14031> mój udział polegał wsparciu w opracowaniu metodologii badania, współredagowaniu artykułu oraz nadzorowaniu pracy zespołu badawczego. Mój udział w powstaniu pracy wynosi 10%.


.....
(podpis)

Informacja o przetwarzaniu danych osobowych

Na podstawie Rozporządzenia Parlamentu Europejskiego i Rady (UE) 2016/679 z dnia 27 kwietnia 2016 r. w sprawie ochrony osób fizycznych w związku z przetwarzaniem danych osobowych i w sprawie swobodnego przepływu takich danych oraz uchylenia dyrektywy 95/46/WE, zwanego dalej „RODO”, informujemy, że:

1. Administratorem Pana/Pani danych osobowych będzie Uniwersytet Mikołaja Kopernika w Toruniu z siedzibą przy ul. Gagarina 11, 87-100 Toruń (dalej: Uczelnia, ADO).
2. Pana/Pani dane osobowe w związku z postępowaniem o nadanie stopnia naukowego będą przetwarzane:
 - a. na podstawie art. 6 ust. 1 lit. c) RODO – obowiązek prawny wynikający z przepisów ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (Dz. U. z 2020, poz. 85);
 - b. na podstawie art. 6 ust. 1 lit. f) RODO – prawnie uzasadniony interes ADO:
 - do ustalenia, obrony lub dochodzenia roszczeń – przez okres do przedawnienia roszczeń, lub przez okres prowadzenia postępowania przez właściwe organy lub sądy w przypadku dochodzenia roszczeń,
 - na potrzeby wewnętrzne tworzenia zestawień, analiz i statystyk – przez okres obowiązywania umowy,
 - na potrzeby prowadzenia ewidencji korespondencji przychodzącej i wychodzącej – wieceżyscie,
 - na potrzeby marketingu produktów i usług ADO.
3. Z zastrzeżeniem przepisów powszechnie obowiązującego prawa przysługują Panu/Pani prawa, które zrealizujemy na wniosek o:
 - a. Żądanie dostępu do danych osobowych oraz prawo ich sprostowania,
 - b. Żądanie usunięcia lub ograniczenia przetwarzania.
4. Przysługuje Panu/Pani również prawo wniesienia sprzeciwu na przetwarzanie danych osobowych.
5. Podanie przez Pana/Panią danych osobowych jest niezbędne do wykonania celu wymienionego w pkt 2 lit. a i b., a brak ich podania uniemożliwi otwarcie i przeprowadzenie postępowania.
6. Przysługuje Panu/Pani prawo wniesienia skargi do Prezesa Urzędu Ochrony Danych Osobowych.
7. Pana/Pani dane osobowe mogą być udostępnione recenzentom lub organom administracji publicznej, sądom, komornikom w zakresie sytuacji przewidzianych w przepisach prawa.
8. Na dzień zbierania Pana/Pani danych osobowych nie planujemy przekazywać ich poza EOG (obejmujący Unię Europejską, Norwegię, Lichtenstein i Islandię), nie wykluczając tego w przyszłości, o czym zostanie Pan/Pani poinformowania ze stosownym wyprzedzeniem.
9. W stosunku do Pana/Pani nie będą prowadzone działania polegające na podejmowaniu decyzji w sposób zautomatyzowany, nie będą one również podlegały zautomatyzowanemu profilowaniu.
10. Jeżeli chce Pan/Pani skontaktować się z Uczelnią w sprawach związanych z przetwarzaniem danych osobowych, w szczególności w związku z wniesieniem wniosku o realizację przysługujących praw prosimy o kontakt pod adresem e-mail: iod@umk.pl lub adresem korespondencyjnym: UMK w Toruniu, ul. Gagarina 11, 87-100 Toruń, z dopiskiem „IOD”, dostępny jest również kontakt telefoniczny: 56 611 27 42.

Toruń, dnia 06.05.2024

Lek. Joanna Stankiewicz

Katedra Pediatrii, Hematologii i Onkologii
Collegium Medicum w Bydgoszczy
Uniwersytet Mikołaja Kopernika w Toruniu

**Rada Dyscypliny Nauk Medycznych
Uniwersytetu Mikołaja Kopernika w
Toruniu**

Oświadczenie o współautorstwie

Niniejszym oświadczam, że w pracy Michalski A, Lis K, Stankiewicz J, Kloska S.M, Dudziński M, Muras-Szwedziak K, Nowicki M, Bazan-Socha S, Dąbrowski M.J, Basak G.W. Supporting the diagnosis of Fabry disease using a natural language processing-based approach. J. Clin. Med. 2023, 12(10), 3599; <https://doi.org/10.3390/jcm12103599> mój udział polegał na wsparciu w opracowaniu metodologii badania, współredagowaniu artykułu oraz analizie dokumentacji medycznej grupy badanej i kontrolnej. Mój udział w powstaniu pracy wynosi 5%.

Joanna Stankiewicz
(podpis)

Informacja o przetwarzaniu danych osobowych

Na podstawie Rozporządzenia Parlamentu Europejskiego i Rady (UE) 2016/679 z dnia 27 kwietnia 2016 r. w sprawie ochrony osób fizycznych w związku z przetwarzaniem danych osobowych i w sprawie swobodnego przepływu takich danych oraz uchylenia dyrektywy 95/46/WE, zwanego dalej „RODO”, informujemy, że:

1. Administratorem Pana/Pani danych osobowych będzie Uniwersytet Mikołaja Kopernika w Toruniu z siedzibą przy ul. Gagarina 11, 87-100 Toruń (dalej: Uczelnia, ADO).
2. Pana/Pani dane osobowe w związku z postępowaniem o nadanie stopnia naukowego będą przetwarzane:
 - a. na podstawie art. 6 ust. 1 lit. c) RODO – obowiązek prawny wynikający z przepisów ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (Dz. U. z 2020, poz. 85);
 - b. na podstawie art. 6 ust. 1 lit. f) RODO – prawnie uzasadniony interes ADO:
 - do ustalenia, obrony lub dochodzenia roszczeń – przez okres do przedawnienia roszczeń, lub przez okres prowadzenia postępowania przez właściwe organy lub sądy w przypadku dochodzenia roszczeń,
 - na potrzeby wewnętrzne tworzenia zestawień, analiz i statystyk – przez okres obowiązywania umowy,
 - na potrzeby prowadzenia ewidencji korespondencji przychodzącej i wychodzącej – wieczyście,
 - na potrzeby marketingu produktów i usług ADO.
3. Z zastrzeżeniem przepisów powszechnie obowiązującego prawa przysługują Panu/Pani prawa, które zrealizujemy na wniosek o:
 - a. Żądanie dostępu do danych osobowych oraz prawo ich sprostowania,
 - b. Żądanie usunięcia lub ograniczenia przetwarzania.
4. Przysługuje Panu/Pani również prawo wniesienia sprzeciwu na przetwarzanie danych osobowych.
5. Podanie przez Pana/Panią danych osobowych jest niezbędne do wykonania celu wymienionego w pkt 2 lit. a i b., a brak ich podania uniemożliwi otwarcie i przeprowadzenie postępowania.
6. Przysługuje Panu/Pani prawo wniesienia skargi do Prezesa Urzędu Ochrony Danych Osobowych.
7. Pana/Pani dane osobowe mogą być udostępnione recenzentom lub organom administracji publicznej, sądom, komornikom w zakresie sytuacji przewidzianych w przepisach prawa.
8. Na dzień zbierania Pana/Pani danych osobowych nie planujemy przekazywać ich poza EOG (obejmujący Unię Europejską, Norwegię, Lichtenstein i Islandię), nie wykluczając tego w przyszłości, o czym zostanie Pan/Pani poinformowany ze stosownym wyprzedzeniem.
9. W stosunku do Pana/Pani nie będą prowadzone działania polegające na podejmowaniu decyzji w sposób zautomatyzowany, nie będą one również podlegały zautomatyzowanemu profilowaniu.
10. Jeżeli chce Pan/Pani skontaktować się z Uczelnią w sprawach związanych z przetwarzaniem danych osobowych, w szczególności w związku z wniesieniem wniosku o realizację przysługujących praw prosimy o kontakt pod adresem e-mail: iod@umk.pl lub adresem korespondencyjnym: UMK w Toruniu, ul. Gagarina 11, 87-100 Toruń, z dopiskiem „IOD”, dostępny jest również kontakt telefoniczny: 56 611 27 42.



Toruń, dnia 06.05.2024

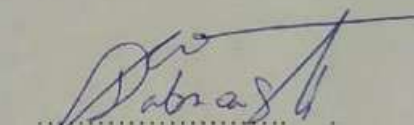
dr Michał J. Dąbrowski

Zespół Biologii Obliczeniowej
Instytut Podstaw Informatyki Polskiej Akademii Nauk

**Rada Dyscypliny Nauk Medycznych
Uniwersytetu Mikołaja Kopernika w
Toruniu**

Oświadczenie o współautorstwie

Niniejszym oświadczam, że w pracy Michalski A, Lis K, Stankiewicz J, Kloska S.M, Dudziński M, Muras-Szwedziak K, Nowicki M, Bazan-Socha S, Dąbrowski M.J, Basak G.W. Supporting the diagnosis of Fabry disease using a natural language processing-based approach. J. Clin. Med. 2023, 12(10), 3599; <https://doi.org/10.3390/jcm12103599> mój udział polegał na wsparciu w opracowaniu metodologii badania, współredagowaniu artykułu oraz nadzorowaniu pracy zespołu badawczego. Mój udział w powstaniu pracy wynosi 9%.


.....
(podpis)

Informacja o przetwarzaniu danych osobowych

Na podstawie Rozporządzenia Parlamentu Europejskiego i Rady (UE) 2016/679 z dnia 27 kwietnia 2016 r. w sprawie ochrony osób fizycznych w związku z przetwarzaniem danych osobowych i w sprawie swobodnego przepływu takich danych oraz uchylenia dyrektywy 95/46/WE, zwanego dalej „RODO”, informujemy, że:

1. Administratorem Pana/Pani danych osobowych będzie Uniwersytet Mikołaja Kopernika w Toruniu z siedzibą przy ul. Gagarina 11, 87-100 Toruń (dalej: Uczelnia, ADO).
2. Pana/Pani dane osobowe w związku z postępowaniem o nadanie stopnia naukowego będą przetwarzane:
 - a. na podstawie art. 6 ust. 1 lit. c) RODO – obowiązek prawny wynikający z przepisów ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (Dz. U. z 2020, poz. 85);
 - b. na podstawie art. 6 ust. 1 lit. f) RODO – prawnie uzasadniony interes ADO:
 - do ustalenia, obrony lub dochodzenia roszczeń – przez okres do przedawnienia roszczeń, lub przez okres prowadzenia postępowania przez właściwe organy lub sądy w przypadku dochodzenia roszczeń,
 - na potrzeby wewnętrzne tworzenia zestawień, analiz i statystyk – przez okres obowiązywania umowy,
 - na potrzeby prowadzenia ewidencji korespondencji przychodzącej i wychodzącej – wiecześnie,
 - na potrzeby marketingu produktów i usług ADO.
3. Z zastrzeżeniem przepisów powszechnie obowiązującego prawa przysługują Panu/Pani prawa, które zrealizujemy na wniosek o:
 - a. Żądanie dostępu do danych osobowych oraz prawo ich sprostowania,
 - b. Żądanie usunięcia lub ograniczenia przetwarzania.
4. Przysługuje Panu/Pani również prawo wniesienia sprzeciwu na przetwarzanie danych osobowych.
5. Podanie przez Pana/Panią danych osobowych jest niezbędne do wykonania celu wymienionego w pkt 2 lit. a i b., a brak ich podania uniemożliwi otwarcie i przeprowadzenie postępowania.
6. Przysługuje Panu/Pani prawo wniesienia skargi do Prezesa Urzędu Ochrony Danych Osobowych.
7. Pana/Pani dane osobowe mogą być udostępnione recenzentom lub organom administracji publicznej, sądom, komornikom w zakresie sytuacji przewidzianych w przepisach prawa.
8. Na dzień zbierania Pana/Pani danych osobowych nie planujemy przekazywać ich poza EOG (obejmujący Unię Europejską, Norwegię, Lichtenstein i Islandię), nie wykluczając tego w przyszłości, o czym zostanie Pan/Pani poinformowany ze stosownym wyprzedzeniem.
9. W stosunku do Pana/Pani nie będą prowadzone działania polegające na podejmowaniu decyzji w sposób zautomatyzowany, nie będą one również podlegały zautomatyzowanemu profilowaniu.
10. Jeżeli chce Pan/Pani skontaktować się z Uczelnią w sprawach związanych z przetwarzaniem danych osobowych, w szczególności w związku z wniesieniem wniosku o realizację przysługujących praw prosimy o kontakt pod adresem e-mail: iod@umk.pl lub adresem korespondencyjnym: UMK w Toruniu, ul. Gagarina 11, 87-100 Toruń, z dopiskiem „IOD”, dostępny jest również kontakt telefoniczny: 56 611 27 42.

Toruń, dnia 06.05.2024


dr n.med. Sylwester M. Kloska

Wydział Medyczny
Politechnika Bydgoska

**Rada Dyscypliny Nauk Medycznych
Uniwersytetu Mikołaja Kopernika w
Toruniu**

Oświadczenie o współautorstwie

Niniejszym oświadczam, że w pracy Michalski A, Lis K, Stankiewicz J, Kloska S.M, Dudziński M, Muras-Szwedziak K, Nowicki M, Bazan-Socha S, Dąbrowski M.J, Basak G.W. Supporting the diagnosis of Fabry disease using a natural language processing-based approach. J. Clin. Med. 2023, 12(10), 3599; <https://doi.org/10.3390/jcm12103599> mój udział polegał na udziale w opracowaniu metodologii badania, współredagowaniu artykułu oraz analizie statystycznej. Mój udział w powstaniu pracy wynosi 5%.


.....
(podpis)

Informacja o przetwarzaniu danych osobowych

Na podstawie Rozporządzenia Parlamentu Europejskiego i Rady (UE) 2016/679 z dnia 27 kwietnia 2016 r. w sprawie ochrony osób fizycznych w związku z przetwarzaniem danych osobowych i w sprawie swobodnego przepływu takich danych oraz uchylenia dyrektywy 95/46/WE, zwanego dalej „RODO”, informujemy, że:

1. Administratorem Pana/Pani danych osobowych będzie Uniwersytet Mikołaja Kopernika w Toruniu z siedzibą przy ul. Gagarina 11, 87-100 Toruń (dalej: Uczelnia, ADO).
2. Pana/Pani dane osobowe w związku z postępowaniem o nadanie stopnia naukowego będą przetwarzane:
 - a. na podstawie art. 6 ust. 1 lit. c) RODO – obowiązek prawny wynikający z przepisów ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (Dz. U. z 2020, poz. 85);
 - b. na podstawie art. 6 ust. 1 lit. f) RODO – prawnie uzasadniony interes ADO:
 - do ustalenia, obrony lub dochodzenia roszczeń – przez okres do przedawnienia roszczeń, lub przez okres prowadzenia postępowania przez właściwe organy lub sądy w przypadku dochodzenia roszczeń,
 - na potrzeby wewnętrzne tworzenia zestawień, analiz i statystyk – przez okres obowiązywania umowy,
 - na potrzeby prowadzenia ewidencji korespondencji przychodzącej i wychodzącej – wieceście,
 - na potrzeby marketingu produktów i usług ADO.
3. Z zastrzeżeniem przepisów powszechnie obowiązującego prawa przysługują Panu/Pani prawa, które zrealizujemy na wniosek o:
 - a. Żądanie dostępu do danych osobowych oraz prawo ich sprostowania,
 - b. Żądanie usunięcia lub ograniczenia przetwarzania.
4. Przysługuje Panu/Pani również prawo wniesienia sprzeciwu na przetwarzanie danych osobowych.
5. Podanie przez Pana/Panią danych osobowych jest niezbędne do wykonania celu wymienionego w pkt 2 lit. a i b., a brak ich podania uniemożliwi otwarcie i przeprowadzenie postępowania.
6. Przysługuje Panu/Pani prawo wniesienia skargi do Prezesa Urzędu Ochrony Danych Osobowych.
7. Pana/Pani dane osobowe mogą być udostępnione recenzentom lub organom administracji publicznej, sądom, komornikom w zakresie sytuacji przewidzianych w przepisach prawa.
8. Na dzień zbierania Pana/Pani danych osobowych nie planujemy przekazywać ich poza EOG (obejmujący Unię Europejską, Norwegię, Lichtenstein i Islandię), nie wykluczając tego w przyszłości, o czym zostanie Pan/Pani poinformowania ze stosownym wyprzedzeniem.
9. W stosunku do Pana/Pani nie będą prowadzone działania polegające na podejmowaniu decyzji w sposób zautomatyzowany, nie będą one również podlegały zautomatyzowanemu profilowaniu.
10. Jeżeli chce Pan/Pani skontaktować się z Uczelnią w sprawach związanych z przetwarzaniem danych osobowych, w szczególności w związku z wniesieniem wniosku o realizację przysługujących praw prosimy o kontakt pod adresem e-mail: iod@umk.pl lub adresem korespondencyjnym: UMK w Toruniu, ul. Gagarina 11, 87-100 Toruń, z dopiskiem „IOD”, dostępny jest również kontakt telefoniczny: 56 611 27 42.

Uniwersytet Mikołaja Kopernika w Toruniu
Collegium Medicum im L. Rydygiera w Bydgoszczy

KOMISJA BIOETYCZNA

Ul. M. Skłodowskiej-Curie 9, 85-094 Bydgoszcz, tel.(052) 585-35-63, fax.(052) 585-38-11

KB 487/2022

Bydgoszcz, 27.09.2022 r.

Działając na podstawie art. 29 ustawy z dnia 5.12.1996 r. o zawodach lekarza i lekarza dentystry Dz.U. z 1997 r. Nr 28 poz. 152 (wraz z późniejszymi zmianami), rozporządzenia Ministra Zdrowia i Opieki Społecznej z dnia 11.05.1999 r. w sprawie szczegółowych zasad powoływania i finansowania oraz trybu działania komisji bioetycznych (Dz.U. Nr 47 poz.480) oraz Zarządzenia Nr 21 Rektora UMK z dnia 4.03.2009 r. z późn. zm. w sprawie powołania oraz zasad działania Komisji Bioetycznej Uniwersytetu Mikołaja Kopernika w Toruniu przy Collegium Medicum im Ludwika Rydygiera w Bydgoszczy oraz zgodnie z zasadami zawartymi w DH i GCP

Komisja Bioetyczna przy UMK w Toruniu, Collegium Medicum w Bydgoszczy

(skład podano w załączeniu), na posiedzeniu w dniu **27.09.2022 r.** przeanalizowała wniosek, który złożył kierownik badania:

prof. dr hab. n. med. Grzegorz W. Basak
Saventic Health sp. z o. o.
ul. Władysława Łokietka 5
87-100 Toruń

z zespołem w składzie:

mgr Adrian A. Michalski, lek. Karol Lis, lek. Joanna Stankiewicz, mgr Sylwester M. Kłoska, dr Michał J. Dąbrowski, prof. dr hab. n. med. Grzegorz W. Basak

w sprawie badania:

„Diagnostyka chorób rzadkich z wykorzystaniem metod sztucznej inteligencji na podstawie analizy danych medycznych pacjentów”.

Po zapoznaniu się ze złożonym wnioskiem i w wyniku przeprowadzonej dyskusji oraz głosowania Komisja podjęła:

Uchwałę o pozytywnym zaopiniowaniu wniosku

w sprawie przeprowadzenia badań w zakresie określonym we wniosku pod warunkiem uzyskania zgody osób badanych na przetwarzanie danych osobowych w celach naukowych, a w przypadku braku takiej zgody, analizowania jedynie danych zanonimizowanych, pozbawionych danych personalnych (zgodnie z RODO).

Zgoda obowiązuje od daty podjęcia uchwały (27.09.2022 r.) do końca 2023 r.

Wydana opinia dotyczy tylko rozpatrywanego wniosku z uwzględnieniem przedstawionego projektu; każda zmiana i modyfikacja wymaga uzyskania odrębnej opinii

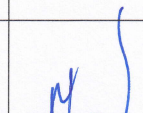
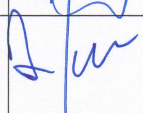
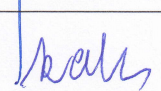
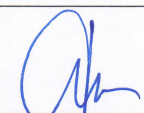
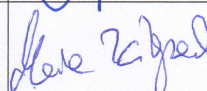
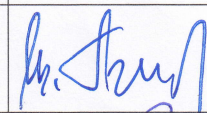
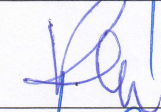
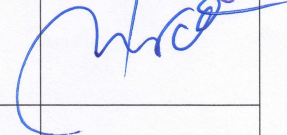
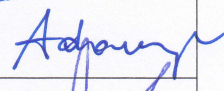
Prof. dr hab. med. Karol Śliwka

Przewodniczący Komisji Bioetycznej

Otrzymuje:

prof. dr hab. n. med. Grzegorz W. Basak
Saventic Health sp. z o. o.
ul. Władysława Łokietka 5
87-100 Toruń

Lista obecności
na posiedzeniu Komisji Bioetycznej
w dniu 27.09.2022 r.

Lp.	Imię i nazwisko	Funkcja/ Specjalizacja	Podpis
1.	Prof. dr hab. med. Karol Śliwka	<i>medycyna sądowa</i>	
2.	Mgr prawa Joanna Połetek-Żygas	<i>prawniczka</i>	
3.	Prof. dr hab. med. Mieczysława Czerwionka-Szaflarska	<i>pediatra, alergologia i gastroenterologia dziecięca</i>	
4.	Prof. dr hab. med. Marek Grabiec	<i>położnictwo, ginekologia onkologiczna</i>	
5.	Prof. dr hab. n med. Maria Kłopočka	<i>choroby wewnętrzne, gastroenterologia</i>	
6.	Prof. dr hab. med. Zbigniew Włodarczyk	<i>chirurgia ogólna, transplantologia kliniczna</i>	
7.	Dr hab. n. med. Maciej Słupski, prof. UMK	<i>chirurgia ogólna, transplantologia kliniczna</i>	
8.	Dr hab. n. med. Katarzyna Sierakowska, prof. UMK	<i>anestezjologia i intensywna terapia</i>	
9.	Ks. dr hab. Wojciech Szukalski, prof. UAM	<i>duchowny</i>	
10.	Dr n. med. Radosława Staszak-Kowalska	<i>pediatria, choroby płuc</i>	
11.	Mgr prawa Patrycja Brzezicka	<i>prawniczka</i>	
12.	Mgr farm. Aleksandra Adamczyk	<i>farmaceutka</i>	
13.	Mgr Lidia Iwińska-Tarczykowska	<i>pielęgniarka</i>	