Magda Konieczna[1]

Akademia Bialska im. Jana Pawła II, Biała Podlaska

https://doi.org/10.29316/9788368103304_4

SYSTEMY KOMPUTEROWEGO WSPOMAGANIA DIAGNOSTYKI OPARTE NA GŁĘBOKIM UCZENIU I TECHNIKACH WYJAŚNIANIA DECYZJI: AKTUALNE TRENDY I KIERUNKI ROZWOJU

COMPUTER-AIDED DIAGNOSTIC SYSTEMS BASED ON DEEP LEARNING AND DECISION EXPLANATION TECHNIQUES: CURRENT TRENDS AND DIRECTIONS OF DEVELOPMENT

Streszczenie

Celem rozdziału jest podsumowanie badań z lat 2022-2025 dotyczących systemów komputerowego wspomagania diagnostyki (CAD, *Computer-Aided Diagnosis*) opartych na głębokim uczeniu (DL, *Deep Learning*) z komponentem interpretowalności (XAI, *Explainable Artificial Intelligence*). Analizie poddano 76 publikacji spełniających kryteria włączenia. Dominującymi zadaniami badawczymi były klasyfikacja (63,2%), segmentacja (15,8%) i detekcja (14,5%), a w mniejszym stopniu podejścia wielozadaniowe.

[1] e-mail: m.konieczna@dyd.akademiabialska.pl

Najczęściej stosowane modalności obejmowały MRI (30,3%), RTG (28,9%) oraz CT (23,7%). W architekturach dominowały klasyczne konwolucyjne sieci neuronowe (CNN, *Convolutional Neural Network*; np. ResNet, VGG, DenseNet), rośnie jednak udział hybrydowych modeli CNN-Transformer, szczególnie w analizach trójwymiarowych (3D). Wszystkie badania uwzględniały komponent XAI. Najczęściej stosowano Grad-CAM (*Gradient-weighted Class Activation Mapping*) i jego warianty, a rzadziej SHAP (*SHapley Additive exPlanations*), LIME (*Local Interpretable Model-agnostic Explanations*) czy *Integrated Gradients*. Narzędzia te poprawiały transparentność modeli, zwiększając zaufanie klinicystów. Typowe metryki skuteczności obejmowały dokładność (Accuracy), pole pod krzywą ROC (AUC, *Area Under the Curve*), miarę F1-score oraz współczynnik podobieństwa Dice'a (Dice, *Dice Similarity Coefficient*); większość modeli osiągała wysokie wyniki (AUC > 0,90, Dice > 0,85). Główne bariery praktycznego wdrożenia to brak standaryzacji metryk, ograniczona liczba badań wieloośrodkowych oraz wysokie koszty obliczeniowe modeli trójwymiarowych (3D) i multimodalnych. Najważniejsze kierunki rozwoju obejmują integrację danych multimodalnych, uczenie federacyjne (FL, *Federated Learning*), standaryzację protokołów walidacyjnych oraz zaawansowane narzędzia XAI. Systemy CAD oparte na DL i XAI mają potencjał, by zwiększyć dokładność oraz transparentność diagnostyki obrazowej, jednak ich pełna implementacja wymaga dalszych badań i walidacji klinicznej.

Słowa kluczowe: głębokie uczenie, sztuczna inteligencja wyjaśnialna, systemy wspomagania diagnostyki, obrazowanie medyczne, modele hybrydowe

Abstract

This chapter summarises research published between 2022 and 2025 on computer-aided diagnosis (CAD) systems based on deep learning (DL) with an explainability component (XAI, Explainable Artificial Intelligence). A total of 76 publications meeting the inclusion criteria were analysed. The dominant research tasks were classification (63.2%), segmentation (15.8%), and detection (14.5%), with a smaller share of multitask approaches. The most frequently used imaging modalities included MRI (30.3%), X-ray (28.9%), and CT (23.7%). Classical convolutional neural networks (CNNs; e.g., ResNet, VGG, DenseNet) have dominated the architecture landscape, although the share of hybrid CNN-Transformer models has been increasing, particularly in three-dimensional (3D) analyses. All studies included an XAI component. The most commonly applied explainability techniques were Grad-CAM (Gradient-weighted Class Activation Mapping) and its variants, while SHAP (Shapley Additive exPlanations), LIME (Local Interpretable Model-agnostic Explanations), and Integrated Gradients were used less frequently. These tools enhanced model transparency and increased clinicians' trust. Typical performance

metrics included Accuracy, the area under the ROC curve (AUC, Area Under the Curve), F1-score, and the Dice coefficient (Dice Similarity Coefficient); most models achieved high performance (AUC > 0.90, Dice > 0.85). The main barriers to practical implementation include the lack of metric standardisation, the limited number of multicenter studies, and the high computational costs of 3D and multimodal models. The most important directions for future research include integrating multimodal data, adopting federated learning (FL), standardising validation protocols, and advancing XAI tools. CAD systems based on DL and XAI have the potential to improve both the accuracy and transparency of medical image diagnosis; however, their full implementation requires further research and clinical validation.

Keywords: deep learning, explainable artificial intelligence, diagnostic support systems, medical imaging, hybrid models

Wstęp

W ostatnich latach głębokie uczenie (ang. deep learning, DL) odegrało kluczową rolę w analizie obrazów medycznych, wspierając proces automatyzacji diagnostyki. Techniki DL umożliwiają wykrywanie, segmentację oraz klasyfikację zmian patologicznych w obrazach pochodzących z różnych modalności, takich jak radiografia (RTG), tomografia komputerowa (CT), rezonans magnetyczny (MRI), ultrasonografia (USG) oraz pozytonowa tomografia emisyjna (PET). Modele oparte na konwolucyjnych sieciach neuronowych (CNN), architekturach typu U-Net oraz nowoczesnych podejściach wykorzystujących mechanizmy uwagi – takich jak Vision Transformers czy Swin Transformer – osiągają wysoką skuteczność w zadaniach diagnostycznych i coraz częściej są integrowane z systemami wspomagania decyzji klinicznych (CAD) (Yamashita i in., 2018; Iqbal i in., 2023; Thanoon i in., 2023; Yeasmin i in., 2024).

Analiza literatury opublikowanej w latach 2022-2025 wskazuje, że modele DL charakteryzują się nie tylko wysoką trafnością diagnostyczną, ale także potencjałem w zakresie zwiększania dostępności usług medycznych oraz skracania czasu diagnostyki, co ma szczególne znaczenie w kontekście rosnącego obciążenia systemów opieki zdrowotnej. Nowoczesne systemy CAD oparte na DL wykazują wysoką skuteczność w detekcji zmian nowotworowych, złamań kostnych, ognisk

niedokrwiennych, zwłóknień oraz patologii naczyniowych – również w przypadku obrazów trudnych do jednoznacznej interpretacji (Shobayo, Saatchi, 2025; Berghout, 2025; Shelmerdine i in., 2022).

Coraz większego znaczenia nabierają również interpretowalność modeli oraz poziom zaufania klinicznego do ich decyzji. Narzędzia z zakresu wyjaśnialnej sztucznej inteligencji (XAI), takie jak Grad-CAM, SHAP czy LIME, odgrywają kluczową rolę w procesie akceptacji technologii DL przez środowiska medyczne, szczególnie w zastosowaniach o podwyższonym ryzyku klinicznym (Hou i in., 2024; Selvaraju i in., 2017; Lundberg, Lee, 2017; Ribeiro i in., 2016).
Celem niniejszego przeglądu jest syntetyczna analiza aktualnego stanu wiedzy na temat zastosowań głębokiego uczenia w analizie obrazów medycznych w latach 2022-2025. W szczególności przeanalizowano:

- nowoczesne architektury i podejścia hybrydowe w dziedzinie DL,
- zastosowania modeli DL w analizie obrazów pochodzących z różnych modalności (RTG, CT, MRI, USG, PET),
- metody oceny skuteczności modeli (m.in. AUC, mAP, IoU, współczynnik Dice),
- narzędzia interpretowalności i ich znaczenie kliniczne,
- aktualne wyzwania oraz potencjalne kierunki dalszego rozwoju technologii.

Metodologia przeglądu

Strategia wyszukiwania i źródła danych

W celu przeprowadzenia przeglądu literatury zastosowano systematyczne wyszukiwanie publikacji naukowych opublikowanych w latach 2022-2025 dotyczących zastosowań głębokiego uczenia w analizie obrazów medycznych. Szczególny nacisk położono na prace zawierające komponent diagnostyczny (CAD), segmentacyjny oraz interpretacyjny (XAI).

Wyszukiwania przeprowadzono w pięciu uznanych bazach danych: PubMed, Scopus, IEEE Xplore, ScienceDirect (Elsevier) oraz SpringerLink. Użyto kombinacji słów kluczowych w języku angielskim, takich jak: *deep learning, medical imaging, computer-aided diagnosis, CNN, XAI, CT, MRI, USG, PET, CAD, segmentation, classification, explainability*. Zakres wyszukiwania ograniczono do

publikacji w języku angielskim, opublikowanych w okresie od 1 stycznia 2022 do 1 lipca 2025. W celu zawężenia wyników zastosowano operatory logiczne AND oraz OR, umożliwiające precyzyjne dopasowanie do założonych kryteriów tematycznych.

Kryteria selekcji publikacji

Do analizy włączono wyłącznie publikacje spełniające następujące warunki:

- artykuły oryginalne (empiryczne) zawierające wyniki badań z wykorzystaniem technik głębokiego uczenia,
- zastosowanie modeli DL w analizie obrazów pochodzących z jednej lub więcej modalności (RTG, CT, MRI, USG, PET),
- obecność komponentu diagnostycznego, klasyfikacyjnego, segmentacyjnego lub wspomagania decyzji klinicznych (CAD),
- uwzględnienie narzędzi interpretowalności (XAI) bądź ich analiza,
- pełny tekst publikacji dostępny w języku angielskim.

Z analizy wykluczono prace spełniające następujące kryteria:

- artykuły przeglądowe, komentarze redakcyjne oraz abstrakty bez dostępu do pełnej treści,
- badania koncentrujące się wyłącznie na danych niemedycznych (np. tekstowych, genomowych),
- publikacje nieuwzględniające zastosowań głębokiego uczenia,
- prace bez powiązania z obrazowaniem medycznym lub diagnostyką kliniczną.

Proces selekcji obejmował niezależną ocenę tytułów i streszczeń, a następnie analizę pełnych tekstów. Publikacje spełniające wszystkie kryteria włączenia zostały uwzględnione w analizie końcowej. Ze względów objętościowych bibliografia końcowa ma charakter reprezentatywny i obejmuje wyłącznie pozycje kluczowe dla głównych nurtów badań oraz metod XAI; pełny korpus 76 publikacji stanowił podstawę analizy syntetycznej autora.

Wyniki przeglądu

Charakterystyka zbioru badań: typy zadań, modalności obrazowania i metody interpretowalności

Analiza 76 publikacji zakwalifikowanych do przeglądu (zgodnie z opisanymi kryteriami selekcji) wykazała, że dominującym kierunkiem badań nad systemami CAD opartymi na głębokim uczeniu pozostaje klasyfikacja obrazów medycznych, 63,2% (48/76). Znacznie rzadziej pojawiają się badania ukierunkowane wyłącznie na segmentację, 15,8% (12/76), lub detekcję, 14,5% (11/76). Zauważalny jest również udział projektów wielozadaniowych, łączących klasyfikację z segmentacją, 3,9% (3/76), lub klasyfikację z detekcją, 2,6% (2/76). Struktura ta sugeruje wyraźny trend rozwoju metod integrujących wiele zadań, co pozostaje zgodne z wnioskami nowszych przeglądów dotyczących CAD i analizy obrazów medycznych (Iqbal i in., 2023; Yeasmin i in., 2024).

Pod względem stosowanych modalności obrazowania największy udział mają badania oparte na rezonansie magnetycznym, 30,3% (23/76), oraz radiografii, 28,9% (22/76). Tomografia komputerowa występuje w 23,7% (18/76) prac, natomiast ultrasonografia w 6,6% (5/76). Pozostałe publikacje wykorzystują dane multimodalne (np. połączenie MRI z CT lub PET), jednak ich liczba pozostaje ograniczona, 3,9% (3/76). Struktura ta odzwierciedla utrwaloną rolę MRI w analizach wymagających wysokiej rozdzielczości i danych trójwymiarowych oraz popularność RTG w diagnostyce chorób płuc i urazów kostnych.

Architektury głębokiego uczenia i podejścia 2D/3D

Analiza 76 publikacji wykazała dominację klasycznych konwolucyjnych sieci neuronowych (CNN), spośród których najczęściej stosowaną pozostaje rodzina ResNet, 44,7% (34/76). Znaczący udział odnotowano także dla architektur VGG, 19,7% (15/76), i DenseNet, 18,4% (14/76). Architektura EfficientNet pojawiła się w 13,2% (10/76) prac, natomiast rzadziej stosowane były sieci Inception/Inc-ResNet, MobileNet i Xception, po 6,6% (5/76). Wskazuje to na stopniowe przechodzenie od prostych konwolucji do bardziej zoptymalizowanych modeli sieci szkieletowych.

Obserwuje się także rosnące zainteresowanie architekturami opartymi na mechanizmach uwagi i transformatorach, które wystąpiły w 10,5% (8/76) badań. W wielu przypadkach transformatory nie były stosowane samodzielnie, lecz w konfiguracjach hybrydowych CNN+Transformer, stanowiących 59,2% (45/76) analizowanych rozwiązań. Trend ten wskazuje na potrzebę integracji tradycyjnych sieci konwolucyjnych z mechanizmami sekwencyjnymi i mechanizmami uwagi. Architektury z rodziny U-Net, w tym warianty z mechanizmami uwagi i w wersji trójwymiarowej, były stosowane w 7,9% (6/76) badań i dominowały w zadaniach segmentacyjnych – od złamań miednicy w CT, przez segmentację guzów wątroby, aż po zmiany w stwardnieniu rozsianym. Klasyczne ramy detekcyjne, takie jak Faster R-CNN, pojawiły się jedynie w 1,3% (1/76) publikacji, co wskazuje na ich ograniczone zastosowanie w badaniach z ostatnich lat.

Z perspektywy wymiaru modelowania zdecydowanie dominują podejścia 2D, 92,1% (70/76), podczas gdy architektury 3D (np. 3D CNN, 3D U-Net) wystąpiły w 7,9% (6/76), głównie w kontekście obrazów CT i MRI. Wysokie koszty obliczeniowe oraz wymagania dotyczące danych ograniczają ich szersze wykorzystanie.

Mechanizmy uwagi i wyjaśnialna sztuczna inteligencja

Wszystkie publikacje zakwalifikowane do niniejszego przeglądu zawierały komponent wyjaśnialnej sztucznej inteligencji (XAI). Najczęściej stosowaną grupę metod stanowiły techniki wizualizacji map aktywacji, w tym Grad-CAM i jego warianty, które dominowały w 65,8% (50/76) badań. Popularność tych metod wynika z ich intuicyjności oraz łatwości implementacji, co znajduje odzwierciedlenie w ich dominacji w analizowanym korpusie (Selvaraju i in., 2017; Hou i in., 2024; Tabela 1).

Metody model-agnostic, takie jak SHAP i LIME, stosowano w 13,2% (10/76) publikacji. SHAP był używany m.in. w diagnostyce chorób neurodegeneracyjnych czy mammografii, natomiast LIME w dermoskopii, histopatologii i diagnostyce chorób płuc. Choć bardziej złożone obliczeniowo, metody te umożliwiają szczegółową analizę istotności cech (Lundberg, Lee, 2017; Ribeiro i in., 2016).

Mechanizmy uwagi integrowano z architekturami segmentacyjnymi (np. U-Net z blokami uwagi) oraz z modelami hybrydowymi CNN-Transformer,

występując w 10,5% (8/76) badań. Rozwiązania te nie tylko poprawiały jakość ekstrakcji cech, lecz także wspierały interpretację kliniczną. Rzadziej stosowano metody alternatywne, takie jak Integrated Gradients czy SmoothGrad, (5,3% (4/76)) oraz Guided Backpropagation. Oferują one wyższą rozdzielczość interpretacji, ale wiążą się z większą złożonością obliczeniową (Sundararajan i in., 2017).

Tabela 1. Częstość stosowania metod XAI w analizowanych publikacjach (n=76)

Metoda XAI	Liczba publikacji (n=76)	Procent (%)
Grad-CAM i warianty	50	65,8
SHAP	6	7,9
LIME	4	5,3
Mechanizmy uwagi	8	10,5
Integrated Gradients / SmoothGrad	4	5,3
Guided Backpropagation	2	2,6

Źródło: opracowanie własne.

Zastosowania kliniczne

Analiza 76 publikacji wykazała, że systemy CAD oparte na głębokim uczeniu znajdują szerokie zastosowanie w różnych modalnościach obrazowania medycznego. Najczęściej wykorzystywane są radiografia (35,5%, 27/76) oraz rezonans magnetyczny (26,3%, 20/76). Tomografia komputerowa obejmuje 21,1% (16/76) prac, natomiast ultrasonografia – 14,5% (11/76). Publikacje multimodalne, łączące MRI z CT lub PET, stanowią jedynie niewielki odsetek – po 1,3% (1/76 każda). Każda z modalności charakteryzuje się odmiennym profilem zastosowań klinicznych, doborem architektur sieciowych oraz zakresem stosowanych metryk skuteczności (Tabela 2). Radiografia (RTG) stanowi jedną z najczęściej analizowanych modalności, szczególnie w diagnostyce chorób płuc, gruźlicy oraz złamań kostnych. W badaniach z lat 2022-2025 dominują modele oparte na konwolucyjnych sieciach neuronowych (ResNet, EfficientNet) oraz na architekturach typu Vision Transformer (ViT), które osiągają wysoką skuteczność klasyfikacyjną (AUC 0,92-0,96, ACC 92-96%).

Tabela 2. Zastosowania systemów CAD z komponentem XAI w różnych modalnościach obrazowania medycznego (2022-2025)

Modalność	Liczba badań (%)	Najczęstsze zadania	Typowe architektury	Zakres metryk
RTG	27 (35,5%)	Klasyfikacja, detekcja (gruźlica, zapalenie płuc, złamania)	CNN (ResNet, EfficientNet), ViT	AUC 0,92-0,96; ACC 92-96%; SEN >90%
MRI	20 (26,3%)	Klasyfikacja, segmentacja (guzy, choroby neurodegeneracyjne)	3D CNN, U-Net, Transformers	AUC 0,95-0,98; ACC >93%; Dice >0,85
CT	16 (21,1%)	Detekcja, segmentacja (złamania, krwotoki, guzy)	3D U-Net, CNN–Transformer, YOLO	SEN 88-95%; IoU >0,80; AUC 0,94-0,97
USG	11 (14,5%)	Klasyfikacja, segmentacja (rak piersi, tarczyca, obrazowanie prenatalne)	CNN, U-Net	AUC 0,90-0,95; ACC 89-94%; Dice >0,82
MRI + CT	1 (1,3%)	Analizy multimodalne (guzy, segmentacja)	Hybrydowe CNN+Transformer	AUC ~0,95; Dice >0,83
MRI + PET	1 (1,3%)	Klasyfikacja i segmentacja (nowotwory, przerzuty)	Hybrydowe (Transformer + CNN)	AUC 0,96-0,98; Dice 0,80-0,86

Źródło: opracowanie własne.

Rezonans magnetyczny (MRI) jest najczęściej wykorzystywaną modalnością w badaniach nad CAD, zwłaszcza w diagnostyce nowotworów mózgu i chorób neurodegeneracyjnych. Zastosowanie znajdują tu głównie sieci 3D CNN, architektury U-Net oraz różne warianty Transformerów, które uzyskują wyniki o wysokiej precyzji (AUC 0,95-0,98; Dice > 0,85).

Tomografia komputerowa (CT) pozostaje podstawowym narzędziem w ocenie urazów czaszkowo-mózgowych, złamań i zmian nowotworowych. Architektury 3D (U-Net, CNN-Transformer, YOLO) osiągają czułość na poziomie 88-95% oraz wysokie wartości IoU (>0,80), co potwierdza ich użyteczność w zadaniach detekcji i segmentacji (Shan i in., 2021).

Ultrasonografia (USG) odgrywa rosnącą rolę w diagnostyce raka piersi, chorób tarczycy oraz w obrazowaniu prenatalnym. Modele oparte na CNN i U-Net zapewniają dokładność w granicach 89-94% oraz współczynniki Dice przekraczające 0,82, co potwierdza potencjał tej modalności mimo ograniczonej liczby badań.

Publikacje multimodalne, choć rzadkie (po 1,3% dla MRI+CT i MRI+PET), wskazują na rosnące zainteresowanie integracją danych obrazowych z różnych źródeł. Modele hybrydowe, zwłaszcza konfiguracje CNN+Transformer, uzyskują w nich bardzo wysoką skuteczność (AUC >0,95; Dice >0,83), co podkreśla znaczenie fuzji danych w złożonych zadaniach diagnostycznych (Lipkova i in., 2022; Li i in., 2024).

Podsumowując, skuteczność systemów CAD różni się w zależności od modalności i rodzaju zadania. RTG i CT dominują w badaniach nad złamaniami oraz chorobami płuc, MRI w diagnostyce nowotworów i chorób ośrodkowego układu nerwowego, USG w diagnostyce onkologicznej i prenatalnej, natomiast PET w ocenie nowotworów i przerzutów. Wspólnym trendem jest rozwój modeli hybrydowych i multimodalnych, które umożliwiają poprawę dokładności diagnostycznej oraz zwiększenie interpretowalności decyzji klinicznych.

Metryki oceny i skuteczność modeli

Analiza 76 publikacji wskazuje, że w badaniach nad systemami CAD opartymi na głębokim uczeniu stosowano szerokie spektrum metryk oceny skuteczności. Najczęściej raportowaną miarą była dokładność (Accuracy), obecną w ponad 60% prac, przy czym typowe wartości mieściły się w zakresie 90-99%, co potwierdza wysoką skuteczność modeli w zadaniach klasyfikacyjnych, zwłaszcza dla obrazów RTG i MRI.

AUC (Area Under the ROC Curve) stosowano w około połowie publikacji, głównie w zadaniach klinicznie istotnych, takich jak diagnostyka krwotoków wewnątrzczaszkowych, nowotworów mózgu czy choroby Alzheimera. Wartości AUC często przekraczały 0,90, a w kilku przypadkach osiągały 0,99-1,00, co wskazuje na wyjątkową odporność modeli na zmiany progu decyzyjnego.

Precision, Recall (czułość i swoistość) były wykorzystywane w około 40% badań, zwłaszcza w kontekście detekcji i klasyfikacji zmian patologicznych.

W połączeniu z F1-score (ok. 35% prac) zapewniały bardziej zrównoważoną ocenę modeli w warunkach niezbalansowanych zbiorów danych, np. w dermoskopii, histopatologii i detekcji przerzutów nowotworowych.

Współczynnik Dice i IoU (Jaccard index) dominowały w badaniach segmentacyjnych (ok. 20% i 10% prac). Typowe wartości Dice mieściły się w zakresie 0,85-0,91, a IoU przekraczało 0,80, co wskazuje na wysoką dokładność modeli przy delineacji struktur anatomicznych w CT i MRI.

Rzadziej stosowane były miary dodatkowe, takie jak Cohen's Kappa, MCC, NPV czy mAP (szczególnie w detekcji obiektów, np. złamań). Choć pojawiały się sporadycznie, pozwalały na dodatkowe ujęcie jakości modeli w kontekście wiarygodności diagnostycznej i równowagi klas.

Podsumowując, badania z lat 2022-2025 charakteryzują się wysokimi wartościami kluczowych wskaźników (AUC > 0,90; Dice > 0,85; Accuracy > 90%). Wskazuje to na duży potencjał systemów CAD opartych na głębokim uczeniu w praktyce klinicznej. Jednocześnie widoczny jest brak pełnej standaryzacji raportowania metryk, co utrudnia bezpośrednie porównywanie wyników między badaniami i instytucjami. Standaryzacja w tym zakresie stanowi istotny kierunek dalszych prac.

Interpretowalność i zaufanie kliniczne

Jednym z kluczowych czynników warunkujących praktyczne wdrożenie systemów CAD do diagnostyki klinicznej jest poziom ich interpretowalności oraz związane z nim zaufanie użytkowników końcowych – lekarzy i radiologów. Mimo wysokiej skuteczności modeli głębokiego uczenia w zadaniach klasyfikacyjnych, detekcyjnych i segmentacyjnych, wciąż istnieje obawa przed ich wykorzystaniem jako „czarnych skrzynek", co rodzi wątpliwości dotyczące bezpieczeństwa i odpowiedzialności klinicznej.

W badaniach przeprowadzonych w ostatnich latach podkreśla się, że wizualizacje generowane przez modele (np. Grad-CAM) mogą zwiększać akceptację lekarzy tylko wtedy, gdy wskazywane obszary są zgodne z wiedzą ekspercką i rzeczywistymi lokalizacjami zmian patologicznych. Jednak liczne prace zwracają uwagę na problem nadmiernego uproszczenia – mapy uwagi bywają zbyt ogólne i nie zawsze odzwierciedlają proces decyzyjny modelu.

Dla praktyki klinicznej istotne jest, aby narzędzia XAI były nie tylko poprawne technicznie, ale również użyteczne diagnostycznie. W analizowanych publikacjach wskazywano, że akceptacja modeli rośnie, gdy interpretacje są prezentowane w sposób zbliżony do rutynowej praktyki (np. nakładanie map uwagi na obrazy RTG lub CT, możliwość interaktywnej analizy istotności cech). Z kolei w zadaniach związanych z chorobami rzadkimi lub nowotworami, gdzie stawką są decyzje o dużym znaczeniu klinicznym, często proponuje się stosowanie metod takich jak SHAP czy LIME, mimo ich złożoności obliczeniowej, ze względu na możliwość bardziej precyzyjnej interpretacji (Hou i in., 2024; Lundberg, Lee, 2017; Ribeiro i in., 2016).

Kwestia zaufania klinicznego ma również istotny wymiar regulacyjny i etyczny, ponieważ we wdrożeniach AI w medycynie rośnie znaczenie transparentności, interpretowalności oraz możliwości audytu modeli medycznych. Oznacza to, że brak interpretowalności może stanowić barierę nie tylko psychologiczną dla lekarzy, ale również organizacyjną i wdrożeniową dla producentów systemów CAD.

Podsumowując, lata 2022-2025 przynoszą wyraźne przesunięcie w stronę badań, które akcentują nie tylko skuteczność diagnostyczną, ale także kliniczną użyteczność i przejrzystość modeli. Implementacja narzędzi XAI – od prostych map aktywacji po zaawansowane metody analizy cech – staje się istotnym warunkiem akceptacji przez użytkowników końcowych i jednym z kluczowych wymiarów jakości współczesnych systemów CAD (Hou i in., 2024).

Dyskusja

Analiza zgromadzonej literatury wskazuje, że dominującą rolę w obszarze interpretowalności systemów CAD opartych na głębokim uczeniu odgrywają techniki wizualizacji map aktywacji, w szczególności Grad-CAM oraz jego warianty. Popularność tych metod wynika z ich intuicyjności i łatwości wizualizacji, które pozwalają klinicystom na szybkie zrozumienie podstaw działania modelu. Niemniej jednak ograniczona zdolność Grad-CAM do szczegółowej i ilościowej analizy cech obrazu istotnie obniża jego przydatność w zastosowaniach wymagających precyzyjnej interpretacji lokalnych struktur anatomicznych.

Równolegle zauważalny jest stopniowy wzrost zainteresowania metodami niezależnymi od architektury modelu (model-agnostic), takimi jak SHAP i LIME, które pozwalają na bardziej szczegółową analizę wpływu poszczególnych cech na decyzję klasyfikacyjną. Przykłady obejmują m.in. badania nad klasyfikacją zmian neurodegeneracyjnych, mammografią oraz cytologią szyjki macicy. W dermatologii i histopatologii metody te wspierają transparentność modeli analizujących obrazy dermoskopowe i preparaty tkankowe. Ich główną wadą pozostają jednak wysokie koszty obliczeniowe i brak dostosowania do pracy w czasie rzeczywistym, co ogranicza ich użyteczność kliniczną.

Szczególne znaczenie zyskują mechanizmy uwagi (attention mechanisms), które stanowią kluczowe ogniwo łączące wzrost skuteczności modeli z poprawą ich interpretowalności. Architektury oparte na Transformerach oraz hybrydowych sieciach CNN-Transformer wykazują wysoki potencjał nie tylko w zakresie zwiększania jakości predykcji, lecz także w generowaniu map uwagi odzwierciedlających obszary istotne z punktu widzenia ekspertów klinicznych.

Pomimo dynamicznego rozwoju badań w zakresie interpretowalności, wciąż niewystarczająco rozwinięta pozostaje problematyka dotycząca modeli 3D oraz podejść multimodalnych (np. CT łączone z danymi klinicznymi, MRI z PET). Rozwiązania te lepiej odzwierciedlają złożoność struktur anatomicznych i procesów patologicznych, a ich wyniki eksperymentalne są obiecujące. Brak systematycznych analiz i walidacji klinicznej w tym obszarze stanowi jednak istotną barierę dla translacji tych modeli do praktyki klinicznej.

Podsumowując, w latach 2022-2025 obserwuje się przesunięcie akcentów badawczych z prostych technik wizualizacji na bardziej zaawansowane metody interpretowalności i modele hybrydowe. Równocześnie utrzymują się istotne ograniczenia badawcze, szczególnie w zakresie zastosowań multimodalnych, modeli 3D oraz standaryzacji metod oceny interpretowalności – które wyznaczają kierunki dalszych badań nad klinicznym wykorzystaniem systemów CAD.

Wyzwania i perspektywy rozwoju

Pomimo dynamicznego rozwoju systemów CAD opartych na głębokim uczeniu i metodach XAI, wciąż istnieje szereg barier ograniczających ich pełne wdrożenie kliniczne. Kluczowe problemy obejmują brak standaryzacji metryk

i protokołów walidacyjnych, ograniczoną dostępność danych wieloośrodkowych, wysokie koszty obliczeniowe modeli 3D oraz trudności z przenoszalnością wyników do realnych warunków diagnostycznych.

Brak jednolitych standardów oceny (np. AUC, Dice, IoU) utrudnia porównywalność wyników i budowę benchmarków, natomiast zróżnicowanie populacyjne i sprzętowe ogranicza uogólnialność opracowanych rozwiązań. Modele 3D i multimodalne, mimo wysokiej dokładności i bogatego kontekstu diagnostycznego, wymagają znacznych zasobów obliczeniowych i infrastrukturalnych, co utrudnia ich wdrożenie w środowiskach klinicznych o ograniczonych możliwościach technicznych.

Dodatkową barierą pozostaje luka pomiędzy skutecznością eksperymentalną a realną implementacją kliniczną – modele osiągające wysokie wyniki w badaniach kontrolowanych często tracą wydajność przy zastosowaniu na nowych danych. W tym kontekście szczególnego znaczenia nabiera potrzeba adaptacyjnych systemów DL, zdolnych do ciągłego uczenia się i walidacji w różnych populacjach pacjentów.

W odpowiedzi na powyższe ograniczenia, kierunki dalszego rozwoju koncentrują się na kilku kluczowych obszarach:

1. Zaawansowane architektury hybrydowe – integrujące CNN z mechanizmami uwagi i Transformerami, szczególnie w analizie obrazów 3D i danych o wysokiej zmienności klinicznej.
2. Integracja multimodalnych źródeł danych – łączenie obrazów CT, MRI, PET i USG z danymi klinicznymi oraz raportami tekstowymi, co pozwala na bardziej kompleksową analizę przypadków.
3. Uczenie federacyjne i prywatność danych – rozwój rozproszonych metod trenowania, umożliwiających współpracę wielu ośrodków bez wymiany danych pacjentów, przy jednoczesnym zwiększeniu ogólności modeli.
4. Personalizacja i modele adaptacyjne – dostosowanie predykcji do indywidualnych cech pacjenta, co może zwiększyć trafność diagnostyczną i ograniczyć liczbę wyników fałszywie dodatnich.
5. Real-time CAD i wdrożenia kliniczne – optymalizacja architektur i infrastruktury obliczeniowej, umożliwiająca analizę obrazów w czasie rzeczywistym, zwłaszcza w diagnostyce urazowej i zabiegowej.

6. Ewolucja narzędzi XAI – przejście od intuicyjnych metod (np. Grad-CAM) ku bardziej formalnym i ilościowym podejściom, takim jak SHAP, LIME czy Integrated Gradients, które mogą zwiększyć zaufanie kliniczne poprzez spójną interpretację wyników.

Wnioski

Analiza 76 publikacji z lat 2022-2025 potwierdza, że systemy CAD oparte na głębokim uczeniu i interpretowalności stanowią jeden z najbardziej dynamicznie rozwijających się kierunków współczesnej diagnostyki obrazowej.

Modele głębokiego uczenia osiągają wysoką skuteczność w klasyfikacji, detekcji i segmentacji obrazów medycznych, często przewyższając tradycyjne metody analizy. Szczególnie obiecujące są architektury hybrydowe oraz modele 3D, które lepiej odzwierciedlają złożoność danych klinicznych i umożliwiają bardziej precyzyjną diagnostykę.

Interpretowalność modeli stanowi warunek akceptacji klinicznej – dominują metody oparte na wizualizacji map aktywacji (Grad-CAM i warianty), które, mimo intuicyjności, wymagają uzupełnienia o bardziej formalne narzędzia (SHAP, LIME, Integrated Gradients).

Kierunki dalszego rozwoju obejmują integrację multimodalnych źródeł danych, wdrażanie uczenia federacyjnego oraz rozwój precyzyjnych metod interpretacji i systematycznej walidacji klinicznej, które stanowią warunek translacji wyników badań do praktyki diagnostycznej.

Podsumowując, systemy CAD oparte na głębokim uczeniu i metodach XAI mają potencjał, aby znacząco poprawić dokładność, przejrzystość i efektywność diagnostyki obrazowej. Kluczowym wyzwaniem pozostaje jednak osiągnięcie równowagi między skutecznością modeli, ich interpretowalnością i zaufaniem klinicznym, co stanowi fundament ich przyszłej implementacji w opiece zdrowotnej.

Literatura

1. Berghout, T. (2025). The Neural Frontier of Future Medical Imaging: A Review of Deep Learning for Brain Tumour Detection. *Journal of Imaging, 11*(1), 2. https://doi.org/10.3390/jimaging11010002

2. Hou, J., Liu, S., Bie, Y., Wang, H., Tan, A., Luo, L., Chen, H. (2024). *Self-eXplainable AI for Medical Image Analysis: A Survey and New Outlooks*. arXiv. https://doi.org/10.48550/arXiv.2410.02331
3. Iqbal, S., Qureshi, A.N., Li, J., Mahmood, T. (2023). On the Analyses of Medical Images Using Traditional Machine Learning Techniques and Convolutional Neural Networks. *Archives of Computational Methods in Engineering, 30*(5), 3173-3233. https://doi.org/10.1007/s11831-023-09899-9
4. Li, Y., Daho, M.E.H., Conze, P.-H., Zeghlache, R., Le Boité, H., Tadayoni, R., Cochener, B., Lamard, M., Quellec, G. (2024). A Review of Deep Learning-Based Information Fusion Techniques for Multimodal Medical Image Classification. *Computers in Biology and Medicine, 177,* 108635. https://doi.org/10.1016/j.compbiomed.2024.108635
5. Lipkova, J., Chen, R.J., Chen, B., Lu, M.Y., Barbieri, M., Shao, D., Vaidya, A.J., Chen, C., Zhuang, L., Williamson, D.F.K., Shaban, M., Chen, T.Y., Mahmood, F. (2022). Artificial Intelligence for Multimodal Data Integration in Oncology. *Cancer Cell, 40*(10), 1095-1110. https://doi.org/10.1016/j.ccell.2022.09.012
6. Lundberg, S.M., Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. W: *Advances in Neural Information Processing Systems 30,* 4765-4774. https://dl.acm.org/doi/10.5555/3295222.3295230
7. Ribeiro, M.T., Singh, S., Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. W: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* 1135-1144. https://doi.org/10.1145/2939672.2939778
8. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D. (2017). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. W: *Proceedings of the IEEE International Conference on Computer Vision (ICCV),* 618-626. https://doi.org/10.1109/ICCV.2017.74
9. Shelmerdine, S.C., White, R.D., Liu, H., Arthurs, O.J., Sebire, N.J. (2022). Artificial Intelligence for Radiological Paediatric Fracture Assessment: A Systematic Review. *Insights into Imaging, 13,* 94. https://doi.org/10.1186/s13244-022-01234-3
10. Shan, W., Guo, J., Mao, X., Zhang, Y., Huang, Y., Wang, S., Li, Z., Meng, X., Zhang, P., Wu, Z., Wang, Q., Liu, Y., He, K., Wang, Y. (2021). Automated Identification of Skull Fractures With Deep Learning: A Comparison Between Object Detection and Segmentation Approach. *Frontiers in Neurology, 12,* 687931. https://doi.org/10.3389/fneur.2021.687931

11. Shobayo, O., Saatchi, R. (2025). Developments in Deep Learning Artificial Neural Network Techniques for Medical Image Analysis and Interpretation. *Diagnostics, 15*(9), 1072. https://doi.org/10.3390/diagnostics15091072
12. Sundararajan, M., Taly, A., Yan, Q. (2017). Axiomatic Attribution for Deep Networks. W: *Proceedings of the 34th International Conference on Machine Learning, 70,* 3319-3328. https://dl.acm.org/doi/10.5555/3305890.3306024
13. Thanoon, M.A., Zulkifley, M.A., Mohd Zainuri, M.A.A., Abdani, S.R. (2023). A Review of Deep Learning Techniques for Lung Cancer Screening and Diagnosis Based on CT Images. *Diagnostics, 13*(16), 2617. https://doi.org/10.3390/diagnostics13162617
14. Yamashita, R., Nishio, M., Do, R.K.G., Togashi, K. (2018). Convolutional Neural Networks: An Overview and Application in Radiology. *Insights into Imaging, 9*(4), 611-629. https://doi.org/10.1007/s13244-018-0639-9
15. Yeasmin, M.N., Al Amin, M., Joti, T.J., Aung, Z., Azim, M.A. (2024). Advances of AI in Image-Based Computer-Aided Diagnosis: A Review. *Array, 23,* 100357. https://doi.org/10.1016/j.array.2024.100357