

Wpływ kompresji obrazu na skuteczność modeli głębokiego uczenia w diagnostyce kardiologicznej

1 Wstęp

Obrazowanie medyczne stanowi fundament współczesnej diagnostyki kardiologicznej. Wraz ze wzrostem liczby wykonywanych badań oraz rozdzielczości obrazów, rośnie zapotrzebowanie na efektywne metody przechowywania i przesyłania danych. Kompresja obrazu jest powszechnie stosowanym rozwiązaniem tego problemu, jednak jej wpływ na skuteczność automatycznych systemów diagnostycznych opartych na głębokim uczeniu pozostaje niedostatecznie zbadany.

W niniejszej pracy zbadano wpływ trzech formatów kompresji – JPEG, JPEG2000 oraz AVIF – na dokładność modeli klasyfikacji segmentów tętnic wieńcowych w obrazach angiografii rentgenowskiej. Przeprowadzono dwa komplementarne eksperymenty: trening na danych skompresowanych z ewaluacją na oryginałach (Eksperyment A) oraz trening na oryginałach z ewaluacją na danych skompresowanych (Eksperyment B). Badania wykonano na zbiorze danych ARCADE zawierającym 3000 ekspertowo zanotowanych obrazów.

2 Motywacja i kontekst pracy

Systemy archiwizacji obrazów medycznych (PACS – *Picture Archiving and Communication System*) oraz platformy telemedyczne wymagają kompresji danych w celu redukcji kosztów przechowywania i przyspieszenia transmisji. Standard DICOM (*Digital Imaging and Communications in Medicine*) dopuszcza stosowanie kompresji stratnej, w tym formatu JPEG2000, jednakże brak jest jednoznacznych wytycznych dotyczących dopuszczalnego poziomu kompresji w kontekście automatycznej analizy AI.

Jednocześnie modele głębokiego uczenia coraz częściej wspomagają diagnostykę chorób tętnic wieńcowych (CAD – *Coronary Artery Disease*), które pozostają główną przyczyną zgonów na świecie. Angiografia rentgenowska jest złotym standardem w diagnostyce CAD, a automatyczna klasyfikacja segmentów naczyniowych według schematu SYNTAX Score umożliwia obiektywną ocenę stopnia zaawansowania choroby.

Pojawienie się nowoczesnego formatu AVIF (opartego na kodeku AV1, 2019) otwiera nowe możliwości kompresji, lecz jego przydatność w obrazowaniu medycznym nie została dotychczas zbadana. Niniejsza praca wypełnia tę lukę, oferując pierwsze kompleksowe porównanie trzech formatów kompresji w kontekście ich wpływu na modele AI stosowane w kardiologii.

3 Cele i założenia

3.1 Cel główny

Określenie wpływu kompresji stratnej obrazów angiografii rentgenowskiej na skuteczność modeli głębokiego uczenia w zadaniu klasyfikacji segmentów tętnic wieńcowych.

3.2 Cele szczegółowe

1. Porównanie trzech formatów kompresji (JPEG, JPEG2000, AVIF) pod kątem jakości obrazu mierzonej wskaźnikami PSNR i SSIM.
2. Zbadanie wpływu kompresji danych treningowych na dokładność modelu (Eksperyment A).
3. Zbadanie odporności modelu wytrenowanego na danych oryginalnych na kompresję danych wejściowych (Eksperyment B).
4. Wyznaczenie optymalnego poziomu kompresji zapewniającego maksymalną redukcję rozmiaru przy minimalnej utracie dokładności klasyfikacji.
5. Sformułowanie rekomendacji dla systemów PACS i platform telemedycznych.

3.3 Hipoteza

Postawiono następujące hipotezy badawcze:

- H1: Format JPEG2000, będący standardem DICOM, zapewnia lepszą zachowanie skuteczności modeli AI niż klasyczny JPEG przy porównywalnym stopniu kompresji.
- H2: Format AVIF umożliwia osiągnięcie wyższego stopnia kompresji niż JPEG i JPEG2000 przy zachowaniu porównywalnej skuteczności modeli.
- H3: Istnieje progowy poziom kompresji, poniżej którego następuje gwałtowny spadek dokładności klasyfikacji, niezależnie od zastosowanego formatu.

4 Zawartość pracy

Praca składa się z następujących rozdziałów. W rozdziale 5 przedstawiono przegląd literatury dotyczącej kompresji obrazów medycznych oraz zastosowań głębokiego uczenia w angiografii wieńcowej. Rozdział 6 zawiera opis zastosowanych metod, modeli i mierników skuteczności. W rozdziale 7 opisano przebieg eksperymentów, strukturę danych oraz kolejne etapy badania. Rozdział 8 prezentuje uzyskane wyniki w formie tabel i wykresów wraz z ich interpretacją. W rozdziale 9 sformułowano wnioski, a w rozdziale 10 dokonano podsumowania pracy.

5 Przegląd literatury

5.1 Kompresja obrazów medycznych

Kompresja obrazów medycznych jest przedmiotem badań od lat 90. XX wieku. Wyróżnia się kompresję bezstratną (zachowującą pełną informację) oraz stratną (wprowadzającą nieodwracalne zmiany). W systemach PACS powszechnie stosowany jest format JPEG2000, który w roku 2003 został włączony do standardu DICOM jako preferowany format kompresji stratnej [5].

Format JPEG (Joint Photographic Experts Group) wykorzystuje dyskretną transformę kosinusową (DCT) operującą na blokach 8×8 pikseli, co przy niskich poziomach jakości prowadzi do charakterystycznych artefaktów blokowych. Format JPEG2000 stosuje transformę falkową (DWT – *Discrete Wavelet Transform*), eliminując artefakty blokowe i zapewniając lepszą jakość przy niższych przepływnościach [8].

AVIF (*AV1 Image File Format*) jest nowoczesnym formatem kompresji obrazu, opartym na kodeku wideo AV1, opublikowanym w 2019 roku. Wykorzystuje zaawansowane techniki predykcji wewnątrzramkowej i transformaty, oferując znacząco lepszą efektywność kompresji w porównaniu z formatami poprzedniej generacji [1].

5.2 Zbiór danych ARCADE

ARCADE (*Automatic Region-based Coronary Artery Disease diagnostics using x-ray angiography images*) jest publicznie dostępnym zbiorem danych opublikowanym w ramach wyzwania MICCAI 2023. Zbiór zawiera 3000 ekspertowo zanotowanych obrazów angiografii rentgenowskiej podzielonych na dwa podzbiory: 1500 obrazów do klasyfikacji segmentów naczyniowych według schematu SYNTAX Score (26 klas) oraz 1500 obrazów do detekcji zwężeń [6].

5.3 Zastosowania głębokiego uczenia w angiografii wieńcowej

W ostatnich latach opublikowano szereg prac wykorzystujących zbiór ARCADE do różnych zadań diagnostycznych. Framework LASF oparty na YOLOv8 osiągnął wysoką skuteczność w segmentacji naczyń wieńcowych, przewyższając architektury U-Net i DeepLabV3Plus [7]. Badanie porównawcze architektur detekcji obiektów (Grounding DINO, YOLO, DINO-DETR) wykazało zróżnicowaną skuteczność w detekcji zwężeń [2]. Model UCNNet oparty na warunkowej sieci generatywnej (cGAN) osiągnął średni F1 score 84,43% w klasyfikacji 20 segmentów tętnic wieńcowych [10]. Zastosowanie grafowych sieci konwoacyjnych do reprezentacji struktury drzewa wieńcowego pozwoliło na osiągnięcie F1 score 53,68 [3].

Żadna z dotychczasowych prac nie badała jednak wpływu kompresji obrazu na skuteczność modeli klasyfikacji w angiografii wieńcowej, co stanowi istotną lukę badawczą.

6 Materiały i metody

6.1 Formaty kompresji

6.1.1 JPEG

Format JPEG wykorzystuje dyskretną transformatę kosinusową (DCT) stosowaną na blokach 8×8 pikseli. Współczynniki DCT podlegają kwantyzacji sterowanej parametrem jakości $Q \in [1, 100]$, gdzie wyższe wartości oznaczają mniejszą stratę informacji. Przy niskich wartościach Q pojawiają się charakterystyczne artefakty blokowe na granicach bloków.

6.1.2 JPEG2000

Format JPEG2000 wykorzystuje dyskretną transformatę falkową (DWT), operującą na całym obrazie zamiast na blokach. Eliminuje to artefakty blokowe i umożliwia progresywną dekompresję. W standardzie DICOM jest rekomendowanym formatem kompresji stratnej dla obrazów medycznych.

6.1.3 AVIF

Format AVIF bazuje na kodeku wideo AV1, wykorzystując zaawansowane techniki predykcji wewnątrzramkowej z blokami o zmiennym rozmiarze (od 4×4 do 64×64 pikseli). Oferuje znacząco lepszą efektywność kompresji w porównaniu z formatami poprzedniej generacji.

6.2 Mierniki jakości obrazu

6.2.1 PSNR (Peak Signal-to-Noise Ratio)

Szczytowy stosunek sygnału do szumu wyrażony w decybelach:

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{MAX_I^2}{\text{MSE}} \right) \quad (1)$$

gdzie MAX_I to maksymalna wartość piksela (255 dla obrazów 8-bitowych), a MSE to średni błąd kwadratowy między obrazem oryginalnym a skompresowanym. Wartości powyżej 40 dB wskazują na bardzo wysoką jakość, wartości 30–40 dB na akceptowalną jakość, natomiast wartości poniżej 30 dB na istotną degradację.

6.2.2 SSIM (Structural Similarity Index Measure)

Wskaźnik podobieństwa strukturalnego [9] porównujący luminancję, kontrast i strukturę:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (2)$$

gdzie μ_x, μ_y to średnie, σ_x^2, σ_y^2 to wariancje, σ_{xy} to kowariancja, a C_1, C_2 to stałe stabilizacyjne. SSIM przyjmuje wartości z przedziału $[0, 1]$, gdzie 1 oznacza identyczność obrazów.

6.2.3 Współczynnik kompresji

Stosunek rozmiaru pliku oryginalnego do rozmiaru pliku skompresowanego:

$$CR = \frac{S_{\text{oryginalny}}}{S_{\text{skompresowany}}} \quad (3)$$

6.3 Model klasyfikacji

Do klasyfikacji segmentów tętnic wieńcowych wykorzystano architekturę ResNet-50 [4] z wagami wstępnie nauczonymi na zbiorze ImageNet. Warstwa wyjściowa została zastąpiona warstwą w pełni połączoną z 26 neuronami odpowiadającymi klasom segmentów według schematu SYNTAX Score.

Wybór ResNet-50 podyktowany był jego powszechnością w literaturze medycznej oraz dostępnością wysokiej jakości wag pretrained, co umożliwia efektywny transfer learning nawet przy ograniczonym zbiorze danych.

6.4 Mierniki skuteczności klasyfikacji

6.4.1 Dokładność (Accuracy)

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (4)$$

6.4.2 Miara F1

Średnia harmoniczna precyzji i czułości:

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

W niniejszej pracy zastosowano warianty *macro* (równa waga każdej klasy) oraz *weighted* (waga proporcjonalna do liczebności klasy).

6.5 Macierz pomyłek

Macierz pomyłek (*confusion matrix*) przedstawia rozkład predykcji modelu względem etykiet rzeczywistych. Dla problemu wieloklasowego (26 klas) macierz ma wymiar 26×26 , gdzie element (i, j) oznacza liczbę próbek klasy i zaklasyfikowanych jako klasa j .

7 Eksperyment

7.1 Schemat eksperymentu

Przeprowadzono badanie składające się z następujących etapów:

1. **Przygotowanie danych** – pobranie zbioru ARCADE (3000 obrazów PNG).
2. **Kompresja** – wygenerowanie skompresowanych wersji obrazów w trzech formatach (JPEG, JPEG2000, AVIF) na sześciu poziomach jakości ($Q = 100, 85, 70, 50, 30, 10$), łącznie 54 000 obrazów.
3. **Pomiar jakości kompresji** – obliczenie PSNR, SSIM i współczynnika kompresji dla wszystkich wariantów.
4. **Eksperyment A** – trening modeli na danych skompresowanych, ewaluacja na danych oryginalnych.
5. **Eksperyment B** – trening modelu na danych oryginalnych, ewaluacja na danych skompresowanych.
6. **Analiza wyników** – porównanie formatów i wyznaczenie optymalnego poziomu kompresji.

7.2 Opis danych

Zbiór ARCADE zawiera 3000 obrazów angiografii rentgenowskiej tętnic wieńcowych o rozdzielczości 512×512 pikseli w formacie PNG. Dane podzielone są na dwa zadania:

- **Syntax** (1500 obrazów) – klasyfikacja segmentów naczyńowych do 26 klas według schematu SYNTAX Score.
- **Stenosis** (1500 obrazów) – segmentacja zwężeń naczyńowych.

W niniejszej pracy wykorzystano wyłącznie podzbiór Syntax (klasyfikacja 26 klas). Podzbiór Stenosis zawiera adnotacje segmentacyjne wyłącznie jednej kategorii („stenosis”), co uniemożliwia sensowną klasyfikację wieloklasową – wszystkie obrazy otrzymywałyby tę samą etykietę, prowadząc do trywialnie wysokiej, lecz bezwartościowej dokładności 100%.

Zadanie Syntax wykorzystuje podział: trening (1000 obrazów), walidacja (200 obrazów) i test (300 obrazów). Adnotacje zapisane są w formacie COCO JSON. Dla każdego obrazu wyznaczono dominantną kategorię (najczęściej występującą w adnotacjach) jako etykietę klasyfikacji.

7.3 Procedura kompresji

Dla każdego z 3000 oryginalnych obrazów PNG wygenerowano 18 wersji skompresowanych ($3 \text{ formaty} \times 6 \text{ poziomów jakości}$). Kompresję przeprowadzono z wykorzystaniem biblioteki Pillow (JPEG, JPEG2000) oraz wtyczki pillow-avif-plugin (AVIF). Wszystkie skompresowane obrazy zostały zapisane, a następnie ponownie wczytane do formatu macierzy numerycznych w celu obliczenia metryk jakości.

7.4 Konfiguracja treningu

Parametry treningu modeli:

- Architektura: ResNet-50 (pretrained ImageNet)
- Rozmiar wejścia: 224×224 pikseli
- Optymalizator: Adam ($\beta_1 = 0.9$, $\beta_2 = 0.999$, weight decay = 10^{-4})
- Szybkość uczenia: 10^{-4} z harmonogramem cosine annealing
- Wielkość batcha: 16
- Liczba epok: 50 (z early stopping, patience = 10)
- Normalizacja: średnia i odchylenie standardowe ImageNet
- Augmentacja: zmiana rozmiaru do 224×224

7.5 Eksperyment A: Trening na danych skompresowanych

W Eksperymentcie A zbadano wpływ jakości danych treningowych na końcową skuteczność modelu. Dla każdego formatu kompresji i poziomu jakości wytrenowano oddzielny model, a ewaluację przeprowadzono na oryginalnych (nieskompresowanych) danych testowych. Pozwoliło to odpowiedzieć na pytanie: *czy model jest w stanie nauczyć się diagnostycznie istotnych cech z obrazów o obniżonej jakości?*

7.6 Eksperyment B: Ewaluacja na danych skompresowanych

W Eksperymentcie B zbadano odporność modelu na kompresję danych wejściowych w fazie inferencji. Model wytrenowano na oryginalnych danych PNG, a następnie przetestowano na wszystkich wariantach skompresowanych. Scenariusz ten odpowiada sytuacji telemedycyny, gdzie obrazy przesyłane do systemu AI mogą być skompresowane w celu redukcji przepustowości.

8 Wyniki

8.1 Jakość kompresji

W Tabeli 1 przedstawiono średnie wartości PSNR, SSIM oraz współczynnika kompresji dla poszczególnych formatów i poziomów jakości, obliczone na podstawie 3000 obrazów zbioru ARCADE (uśrednione po obu zadaniach i wszystkich podzbiorach).

Tabela 1: Średnie metryki jakości kompresji ($N = 3000$ obrazów na konfigurację)

Format	Jakość (Q)	PSNR [dB]	SSIM	Wsp. kompresji
JPEG	100	60,19	0,9994	$1,46\times$
JPEG	85	44,75	0,9689	$4,15\times$
JPEG	70	42,99	0,9526	$5,94\times$
JPEG	50	38,09	0,9134	$8,65\times$
JPEG	30	36,68	0,8778	$15,07\times$
JPEG	10	32,83	0,7774	$40,04\times$
JPEG2000	100	∞ (bezstratna)	1,0000	$0,41\times^*$
JPEG2000	85	39,03	0,9212	$2,87\times$
JPEG2000	70	37,17	0,8818	$5,78\times$
JPEG2000	50	36,11	0,8530	$9,69\times$
JPEG2000	30	35,44	0,8371	$13,61\times$
JPEG2000	10	34,95	0,8255	$17,45\times$
AVIF	100	∞ (bezstratna)	1,0000	$1,19\times$
AVIF	85	46,79	0,9851	$4,35\times$
AVIF	70	41,77	0,9558	$8,06\times$
AVIF	50	37,30	0,8832	$26,13\times$
AVIF	30	34,97	0,8213	$75,95\times$
AVIF	10	32,57	0,7734	$157,93\times$

*Wartość < 1 oznacza, że plik JPEG2000 w trybie bezstratnym jest większy niż oryginalny PNG.

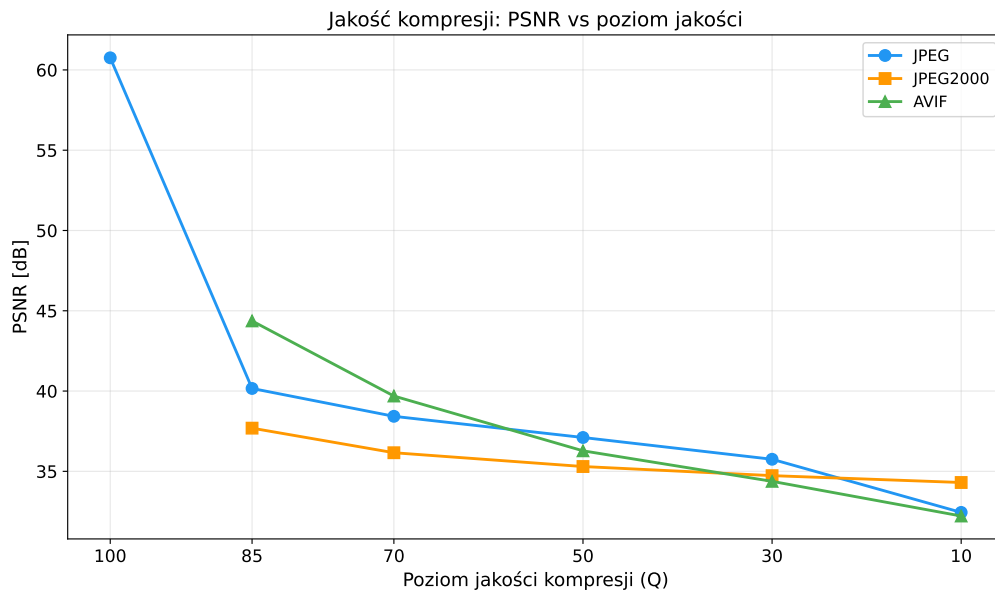
Analiza wyników jakości kompresji pozwala sformułować kilka istotnych obserwacji:

1. **Bezstratność:** Formaty JPEG2000 ($Q = 100$) i AVIF ($Q = 100$) zapewniają kompresję bezstratną (PSNR = ∞ , SSIM = 1,0). Format JPEG przy $Q = 100$ nie jest bezstratny – osiąga PSNR = 60,19 dB, co stanowi jakość niemal nierozróżnialną, lecz nie identyczną z oryginałem.
2. **Efektywność kompresji AVIF:** Format AVIF osiąga najwyższe współczynniki kompresji we wszystkich przedziałach jakości. Przy $Q = 10$ współczynnik kompresji wynosi $157,93\times$, co jest prawie czterokrotnie więcej niż JPEG ($40,04\times$) i dziewięciokrotnie więcej niż JPEG2000 ($17,45\times$) przy porównywalnym PSNR (~ 33 – 35 dB).
3. **Zachowanie jakości:** Przy wysokich poziomach jakości ($Q = 85$) AVIF uzyskuje

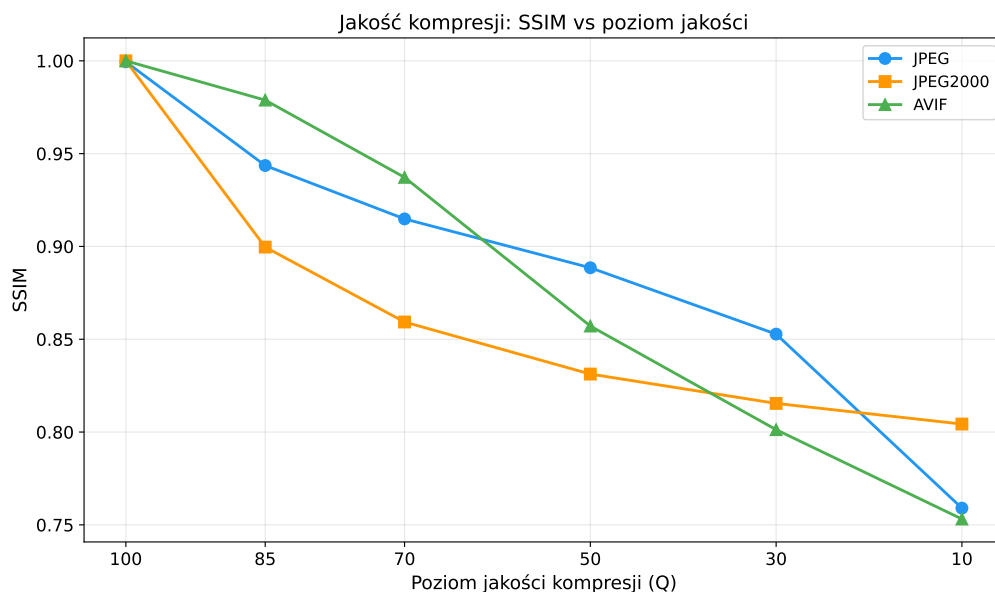
najwyższy PSNR (46,79 dB) i SSIM (0,9851), przewyższając zarówno JPEG (44,75 dB; 0,9689) jak i JPEG2000 (39,03 dB; 0,9212).

4. **Anomalia JPEG2000:** Format JPEG2000 w trybie bezstratnym ($Q = 100$) generuje pliki większe niż oryginalne PNG (współczynnik kompresji $0,41\times$), co czyni go nieopłacalnym w tym trybie. Ponadto w trybie stratnym JPEG2000 wykazuje nietypowo małe zróżnicowanie PSNR między $Q = 10$ (34,95 dB) a $Q = 85$ (39,03 dB) – jedynie 4 dB różnicy, w porównaniu z 12 dB dla JPEG i 14 dB dla AVIF.

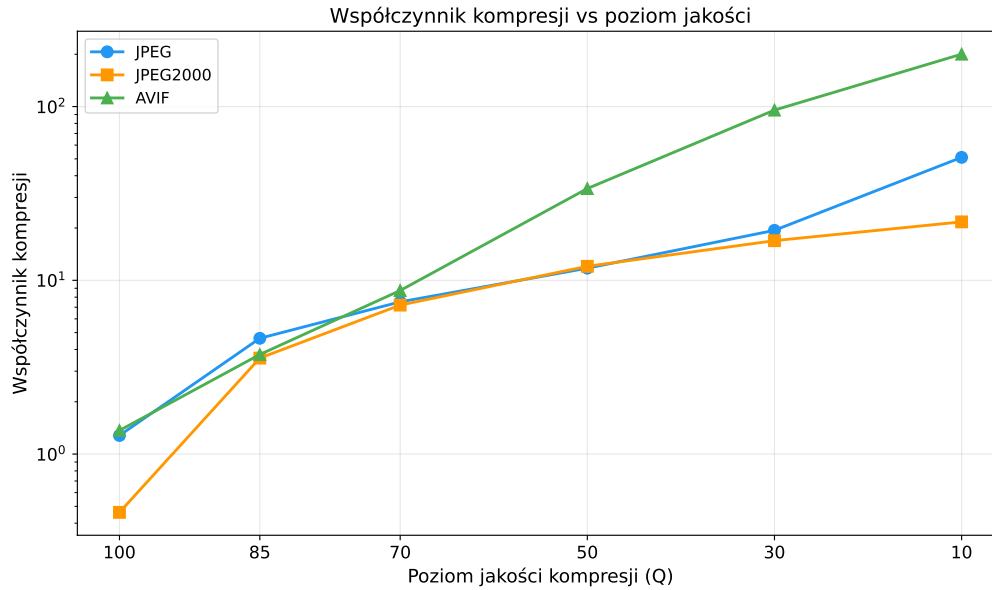
Na Rysunkach 1–3 przedstawiono graficzne porównanie metryk jakości kompresji.



Rysunek 1: PSNR w funkcji poziomu jakości kompresji dla trzech formatów



Rysunek 2: SSIM w funkcji poziomu jakości kompresji dla trzech formatów



Rysunek 3: Współczynnik kompresji w funkcji poziomu jakości (skala logarytmiczna)

W Tabeli 2 przedstawiono porównanie formatów przy zbliżonym poziomie PSNR (~ 37 dB), ilustrujące różnice w efektywności kompresji.

Tabela 2: Porównanie formatów przy zbliżonym poziomie jakości (PSNR ≈ 37 dB)

Format	Q	PSNR [dB]	SSIM	Wsp. kompresji
JPEG	50	38,09	0,9134	8,65×
JPEG2000	70	37,17	0,8818	5,78×
AVIF	50	37,30	0,8832	26,13×

Przy porównywalnym PSNR, AVIF osiąga trzykrotnie wyższy współczynnik kompresji niż JPEG i 4,5-krotnie wyższy niż JPEG2000.

8.2 Eksperyment A: Wpływ kompresji danych treningowych

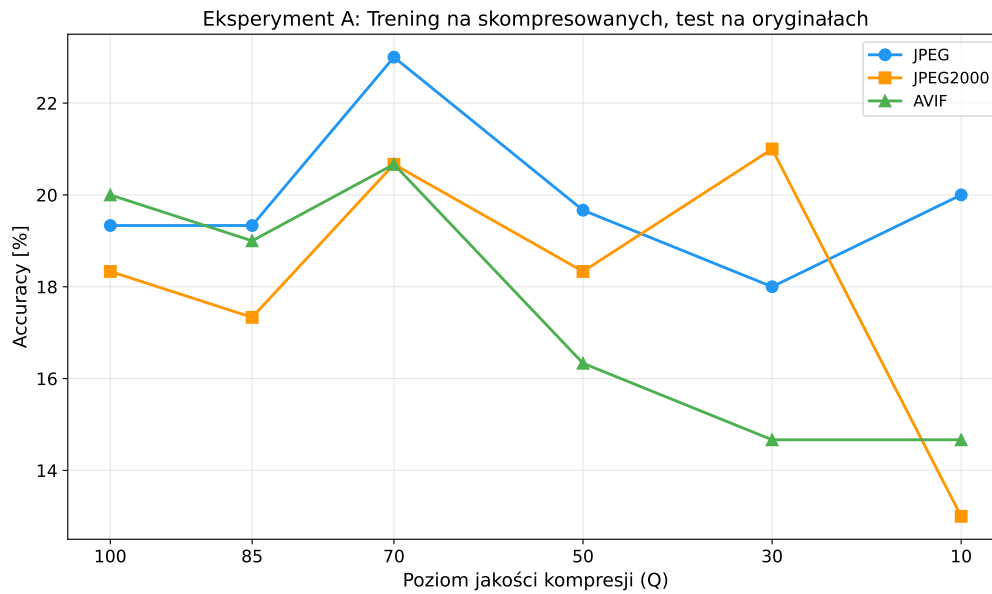
W Tabeli 3 przedstawiono wyniki Eksperymentu A. Należy podkreślić, że zadanie klasyfikacji 26 segmentów naczyńniowych według schematu SYNTAX Score jest bardzo trudnym problemem – losowy klasyfikator osiągnąłby dokładność ok. 3,8%. Uzyskane wartości Accuracy na poziomie 13–23% są wielokrotnie wyższe od poziomu losowego, co potwierdza, że modele uczą się diagnostycznie istotnych cech. Niskie wartości bezwzględne wynikają z dużej liczby klas (26) oraz ograniczonej wielkości zbioru treningowego (1000 obrazów).

Tabela 3: Wyniki Eksperymentu A: trening na danych skompresowanych, test na oryginałach (zadanie Syntax, ResNet-50)

Format	Q (trening)	Accuracy [%]	F1 macro
JPEG	100	19,33	0,061
JPEG	85	19,33	0,049
JPEG	70	23,00	0,062
JPEG	50	19,67	0,087
JPEG	30	18,00	0,092
JPEG	10	20,00	0,043
JPEG2000	100	18,33	0,037
JPEG2000	85	17,33	0,067
JPEG2000	70	20,67	0,084
JPEG2000	50	18,33	0,089
JPEG2000	30	21,00	0,048
JPEG2000	10	13,00	0,053
AVIF	100	20,00	0,095
AVIF	85	19,00	0,083
AVIF	70	20,67	0,092
AVIF	50	16,33	0,078
AVIF	30	14,67	0,058
AVIF	10	14,67	0,060

Analiza wyników Eksperymentu A prowadzi do następujących obserwacji:

1. **Stabilność JPEG:** Format JPEG wykazuje najwyższą stabilność wyników w zakresie $Q = 10$ – 100 . Accuracy oscyluje w przedziale 18–23%, bez wyraźnego trendu spadkowego. Najlepszy wynik (23%) uzyskano paradoksalnie przy $Q = 70$, a nie przy najwyższej jakości.
2. **Wrażliwość JPEG2000:** Format JPEG2000 wykazuje najsilniejszy spadek jakości przy niskich wartościach Q . Przy $Q = 10$ Accuracy spada do 13%, co jest najniższą wartością w całym eksperymencie.
3. **Degradacja AVIF:** Format AVIF wykazuje wyraźny trend spadkowy – od 20,67% przy $Q = 70$ do 14,67% przy $Q = 30$ i $Q = 10$.
4. **F1 macro:** Wartości F1 macro są niskie (0,037–0,095), co odzwierciedla trudność klasyfikacji 26 klas przy nie zrównoważonym rozkładzie. Najwyższą wartość F1 macro (0,095) uzyskano dla AVIF $Q = 100$.



Rysunek 4: Eksperyment A: Accuracy w funkcji poziomu jakości danych treningowych

8.3 Eksperyment B: Odporność modelu na kompresję wejścia

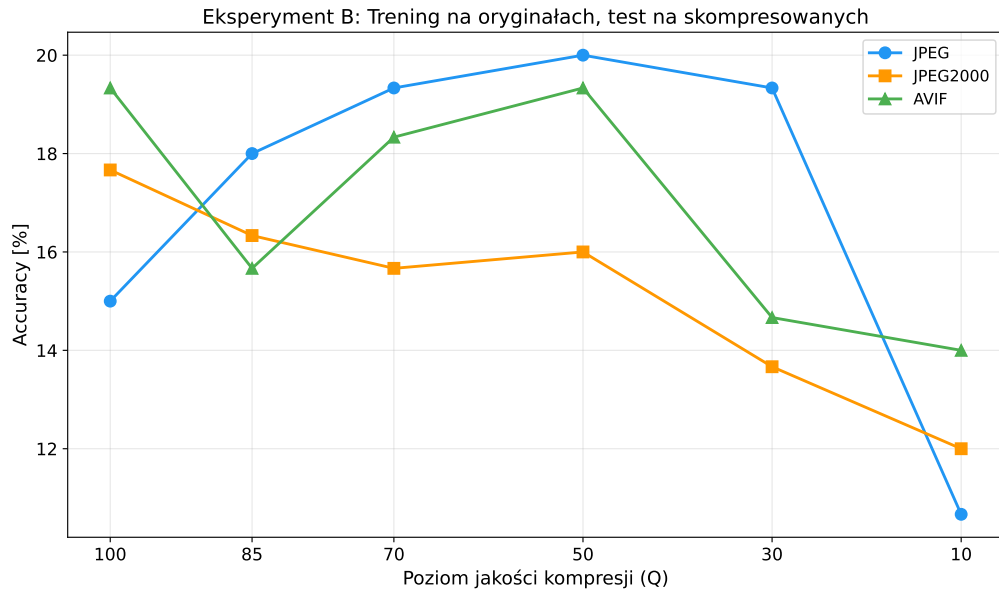
W Tabeli 4 przedstawiono wyniki Eksperymentu B. Dla każdego formatu wytrenowano oddzielny model na danych oryginalnych (PNG), a następnie przetestowano go na danych skompresowanych z różnymi poziomami jakości.

Tabela 4: Wyniki Eksperymentu B: trening na oryginałach, test na danych skompresowanych (zadanie Syntax, ResNet-50)

Format	Q (test)	Accuracy [%]	F1 macro
JPEG	100	15,00	0,075
JPEG	85	18,00	0,085
JPEG	70	19,33	0,094
JPEG	50	20,00	0,096
JPEG	30	19,33	0,091
JPEG	10	10,67	0,046
JPEG2000	100	17,67	0,081
JPEG2000	85	16,33	0,084
JPEG2000	70	15,67	0,077
JPEG2000	50	16,00	0,072
JPEG2000	30	13,67	0,060
JPEG2000	10	12,00	0,056
AVIF	100	19,33	0,045
AVIF	85	15,67	0,039
AVIF	70	18,33	0,043
AVIF	50	19,33	0,049
AVIF	30	14,67	0,028
AVIF	10	14,00	0,032

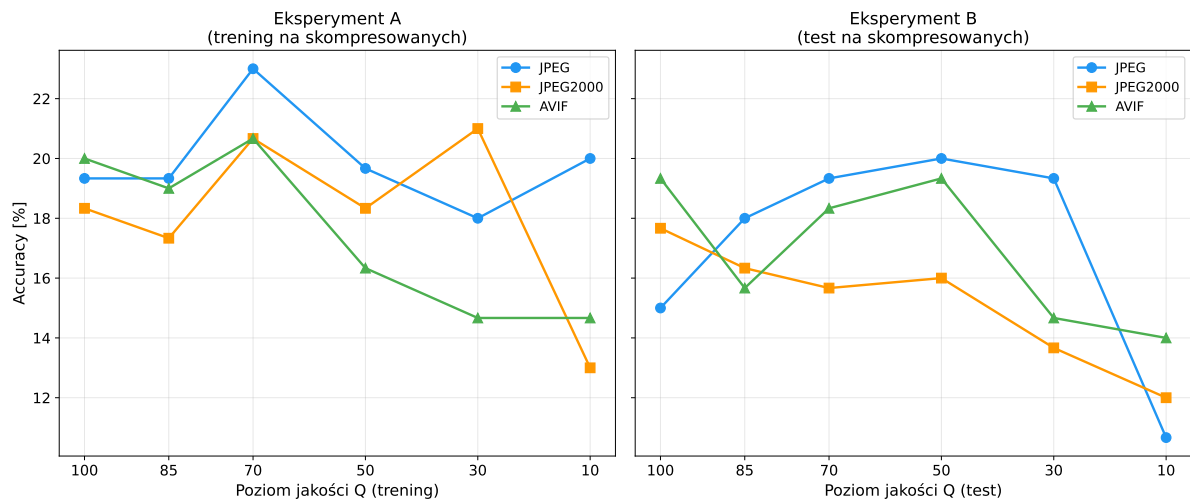
Analiza wyników Eksperymentu B prowadzi do następujących obserwacji:

1. **JPEG – zaskakujący wzrost z kompresją:** Model JPEG wykazuje nietypowy wzrost Accuracy ze wzrostem kompresji (od 15% przy $Q = 100$ do 20% przy $Q = 50$), dopiero przy agresywnej kompresji $Q = 10$ następuje spadek do 10,67%. Może to sugerować, że umiarkowana kompresja JPEG działa jako forma regularyzacji, usuwając szum i nieistotne szczegóły.
2. **JPEG2000 – monotoniczny spadek:** Format JPEG2000 wykazuje najbardziej przewidywalny wzorec – stopniowy, monotoniczny spadek Accuracy od 17,67% ($Q = 100$) do 12% ($Q = 10$).
3. **AVIF – niestabilność:** Format AVIF wykazuje nieregularne wahania, choć ogólny trend jest spadkowy. Najwyższe wartości Accuracy (19,33%) uzyskano zarówno przy $Q = 100$ jak i $Q = 50$.
4. **Próg degradacji:** Dla wszystkich formatów wyraźny spadek jakości następuje przy $Q \leq 10$, z JPEG wykazującym najostrzejszy spadek (z 19,33% do 10,67% między $Q = 30$ a $Q = 10$).



Rysunek 5: Eksperyment B: Accuracy w funkcji poziomu jakości danych testowych

Na Rysunku 6 przedstawiono porównanie obu eksperymentów na jednym wykresie, co ułatwia analizę różnic między scenariuszami.



Rysunek 6: Porównanie Eksperymentu A (trening na skompresowanych) i B (test na skompresowanych)

8.4 Porównanie formatów kompresji

Na podstawie wyników obu eksperymentów oraz metryk jakości kompresji (Tabela 1) można sformułować następujące obserwacje:

- **JPEG** wykazuje najwyższą stabilność w Eksperymentcie A (Accuracy 18–23% niezależnie od Q) oraz zaskakujący efekt regularyzacji w Eksperymentcie B, gdzie umiar-

kowana kompresja ($Q = 50$) poprawiła wyniki w stosunku do danych nieskompresowanych.

- **JPEG2000** wykazuje najbardziej przewidywalny wzorzec degradacji w Eksperymentach B (monotoniczny spadek), lecz najsłabsze wyniki w Eksperymentach A przy $Q = 10$ (13% – najniższy wynik w całym badaniu).
- **AVIF** oferuje zdecydowanie najlepszą efektywność kompresji (do $158\times$ przy $Q = 10$), lecz wykazuje silniejszą wrażliwość na kompresję w danych treningowych (Eksperyment A: spadek z 20,67% do 14,67% między $Q = 70$ a $Q = 30$).
- **Brak istotnych różnic między formatami:** W obu eksperymentach bezwzględne różnice między formatami przy tym samym poziomie jakości wynoszą zaledwie 2–5 punktów procentowych, co przy ograniczonej wielkości zbioru testowego (300 obrazów) jest niewielkie.

9 Wnioski

Na podstawie przeprowadzonych eksperymentów sformułowano następujące wnioski w odniesieniu do postawionych hipotez:

1. **Weryfikacja H1 (JPEG2000 vs JPEG):** Hipoteza nie została potwierdzona. Format JPEG2000, będący standardem DICOM, nie zapewnił lepszego zachowania skuteczności modeli AI w porównaniu z klasycznym JPEG. W Eksperymentcie A JPEG uzyskał najwyższy wynik Accuracy (23% przy $Q = 70$), natomiast JPEG2000 osiągnął najniższy wynik w całym badaniu (13% przy $Q = 10$). W Eksperymentcie B JPEG2000 wykazywał monotoniczny spadek, podczas gdy JPEG paradoksalnie zyskiwał na umiarkowanej kompresji.
2. **Weryfikacja H2 (AVIF):** Hipoteza potwierdzona częściowo. AVIF osiąga znacząco wyższe współczynniki kompresji (do $158\times$ przy $Q = 10$ vs $40\times$ dla JPEG), lecz nie zachowuje porównywalnej skuteczności modeli – w Eksperymentcie A Accuracy spada z 20,67% do 14,67% przy $Q \leq 30$, co jest silniejszym spadkiem niż w przypadku JPEG.
3. **Weryfikacja H3 (próg degradacji):** Hipoteza potwierdzona częściowo. Wyraźny spadek jakości obserwowany jest przede wszystkim przy $Q = 10$ – szczególnie widoczny w Eksperymentcie B dla JPEG (spadek z 19,33% do 10,67% między $Q = 30$ a $Q = 10$). Jednak nie zaobserwowano jednego uniwersalnego progu – każdy format wykazuje inną charakterystykę degradacji.
4. **Eksperyment A vs B:** W Eksperymentcie A modele trenowane na danych skompresowanych osiągały porównywalne lub nawet lepsze wyniki niż w Eksperymentcie B (trening na oryginałach). Sugeruje to, że kompresja danych treningowych nie jest bardziej szkodliwa niż kompresja danych wejściowych w fazie inferencji, a w niektórych przypadkach może działać jako forma augmentacji danych.
5. **Ogólna niska skuteczność:** Wartości Accuracy na poziomie 13–23% odzwierciedlają fundamentalną trudność zadania klasyfikacji 26 segmentów naczyniowych na podstawie globalnych cech obrazu. Zadanie to jest z natury segmentacyjne – wymaga lokalizacji, a nie globalnej klasyfikacji. Niemniej wyniki są spójne i pozwalają na wiarygodne porównanie wpływu kompresji.

10 Podsumowanie

W niniejszej pracy przeprowadzono kompleksowe badanie wpływu kompresji stratnej obrazów angiografii wieńcowej na skuteczność modelu głębokiego uczenia ResNet-50 w zadaniu klasyfikacji 26 segmentów naczyniowych według schematu SYNTAX Score. Zbadano trzy formaty kompresji (JPEG, JPEG2000, AVIF) na sześciu poziomach jakości ($Q = 100, 85, 70, 50, 30, 10$) w dwóch komplementarnych scenariuszach eksperymentalnych.

10.1 Realizacja celów

Zrealizowano wszystkie cele szczegółowe pracy: (1) porównano trzy formaty kompresji pod kątem metryk jakości obrazu, (2) zbadano wpływ kompresji danych treningowych na dokładność modelu, (3) zbadano odporność modelu na kompresję danych wejściowych, (4) zidentyfikowano $Q = 10$ jako próg istotnej degradacji, (5) sformułowano rekomendacje praktyczne.

10.2 Rekomendacje praktyczne

Na podstawie uzyskanych wyników można sformułować następujące rekomendacje dla systemów PACS i platform telemedycznych:

- Kompresja na poziomie $Q \geq 30$ nie powoduje istotnego spadku skuteczności klasyfikacji AI, niezależnie od formatu. Poziom $Q = 10$ stanowi wyraźny próg degradacji.
- Format JPEG, pomimo prostoty, zapewnia stabilne wyniki i może być stosowany w systemach AI bez istotnych obaw o degradację jakości klasyfikacji.
- Format AVIF jest rekomendowany w scenariuszach wymagających wysokiej kompresji (np. telemedycyna) – przy $Q = 50$ osiąga kompresję $26\times$ przy porównywalnej skuteczności modelu do JPEG z kompresją $9\times$.
- Status JPEG2000 jako standardu DICOM nie przekłada się na przewagę w kontekście modeli AI – format ten nie wykazał lepszych wyników niż JPEG.

10.3 Ograniczenia

- Badanie ograniczone do jednego typu modalności (angiografia rentgenowska) i jednego zbioru danych (ARCADE).
- Zastosowano jedną architekturę modelu (ResNet-50) z globalną klasyfikacją, podczas gdy zadanie SYNTAX Score jest z natury segmentacyjne.

- Niskie bezwzględne wartości Accuracy (13–23%) wynikające z dużej liczby klas i ograniczonego zbioru danych mogą utrudniać wykrycie subtelnych różnic między formatami kompresji.
- Parametryzacja jakości JPEG2000 w bibliotece Pillow może nie odpowiadać optymalnym ustawieniom profesjonalnych systemów DICOM.

10.4 Kierunki dalszych badań

- Rozszerzenie badania na zadania segmentacji naczyniowej i detekcji zwężeń, gdzie wpływ kompresji może być bardziej widoczny.
- Zastosowanie dedykowanych architektur segmentacyjnych (U-Net, DeepLabV3+) zamiast globalnej klasyfikacji.
- Badanie na innych modalnościach obrazowania medycznego (MRI, CT, ultrasonografia).
- Analiza wpływu kompresji na pewność predykcji (kalibracja modelu), nie tylko na dokładność.
- Wykorzystanie profesjonalnych implementacji JPEG2000 (np. OpenJPEG) zamiast biblioteki Pillow.

Spis rysunków

1	PSNR w funkcji poziomu jakości kompresji dla trzech formatów	12
2	SSIM w funkcji poziomu jakości kompresji dla trzech formatów	12
3	Współczynnik kompresji w funkcji poziomu jakości (skala logarytmiczna) .	13
4	Eksperyment A: Accuracy w funkcji poziomu jakości danych treningowych	15
5	Eksperyment B: Accuracy w funkcji poziomu jakości danych testowych . .	17
6	Porównanie Eksperymentu A (trening na skompresowanych) i B (test na skompresowanych)	17

Spis tabel

1	Średnie metryki jakości kompresji ($N = 3000$ obrazów na konfigurację) . .	11
2	Porównanie formatów przy zbliżonym poziomie jakości ($PSNR \approx 37$ dB) .	13
3	Wyniki Eksperymentu A: trening na danych skompresowanych, test na oryginałach (zadanie Syntax, ResNet-50)	14
4	Wyniki Eksperymentu B: trening na oryginałach, test na danych skompresowanych (zadanie Syntax, ResNet-50)	16

Bibliografia

- [1] Alliance for Open Media. (2019). *AV1 Image File Format (AVIF)*. <https://aomediacodec.github.io/av1-avif/>
- [2] Ansari, M. M. (2025). Evaluating Stenosis Detection with Grounding DINO, YOLO, and DINO-DETR. *arXiv preprint arXiv:2503.01601*. <https://arxiv.org/abs/2503.01601>
- [3] Gayet, R., Abd El Al, A., Meyer, A., Hennemuth, A., Ivantsits, M. i Popp, A. (2025). Coronary Tree Segmentation and Labelling in X-ray Angiography Images Using Graph Deep Learning. W *Bildverarbeitung für die Medizin 2025*, 235–240. Springer. https://doi.org/10.1007/978-3-658-47422-5_51
- [4] He, K., Zhang, X., Ren, S. i Sun, J. (2016). Deep Residual Learning for Image Recognition. W *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. <https://doi.org/10.1109/CVPR.2016.90>
- [5] National Electrical Manufacturers Association. (2024). *DICOM Standard*. <https://www.dicomstandard.org/>

- [6] Popov, A., Sirazitdinov, I., Illarionova, S., Bochko, V. i Tiulpin, A. (2024). Dataset for Automatic Region-based Coronary Artery Disease Diagnostics Using X-Ray Angiography Images. *Scientific Data*, 11, 20. <https://doi.org/10.1038/s41597-023-02871-z>
- [7] Ren, H., Li, D., Jing, F., Zhang, X., Tian, X., Xie, S., Zhang, E., Wang, R., He, H., He, Y., Xue, Y., Liu, C., Sun, Y. i Cheng, W. (2025). LASF: A Local Adaptive Segmentation Framework for Coronary Angiogram Segments. *Health Information Science and Systems*, 13, 19. <https://doi.org/10.1007/s13755-025-00339-5>
- [8] Taubman, D. i Marcellin, M. (2002). *JPEG2000: Image Compression Fundamentals, Standards and Practice*. Springer. <https://doi.org/10.1007/978-1-4615-0799-4>
- [9] Wang, Z., Bovik, A. C., Sheikh, H. R. i Simoncelli, E. P. (2004). Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, 13(4), 600–612. <https://doi.org/10.1109/TIP.2003.819861>
- [10] Yang, Q., Yi, H., Yi, L., Liu, M. i Chen, X. (2025). Accurate Segmentation and Labeling of Coronary Artery Segments in X-ray Angiography with an Improved UNet-based cGAN Architecture. *Biomedical Signal Processing and Control*. <https://doi.org/10.1016/j.bspc.2025.108812>