

# AUTORZY:

Adam Wieliczko

Wojciech Szlosek

Konrad Nowak

# OPIS I CEL MODELU

Techniki identyfikacji opierają się na m.in. odcisku palca lub DNA

W przypadku katastrof naturalnych takich jak trzęsienia ziemi, powodzie, pożary, a także niektórych przypadkach kryminalnych, cechy takie jak ludzka twarz mogą ulec poważnym uszkodzeniom lub deformacjom, co utrudnia identyfikację.

Autorzy proponują badanie odcisku zębów i identyfikowanie za jej pomocą osoby, gdyż charakterystyczne cechy szkliva pozostają niezmiennie, nawet w niezwykle wysokich temperaturach.



DNA  
recognition



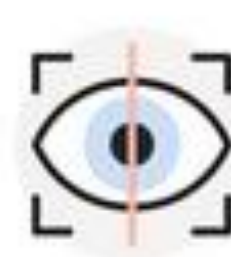
Facial  
recognition



Fingerprint  
recognition



Hand geometry  
recognition



Iris  
recognition

# LIMITACJE

- Pozyskiwanie danych jest wyzwaniem. Promieniowanie rentgenowskie sprawia, że niemożliwe jest pobieranie wielu próbek od jednej osoby. Ponadto, większość danych uzyskiwanych z szpitali składa się z obrazów chorych zębów, które mogą nieodpowiednio odzwierciedlać ogólny stan zdrowia jamy ustnej pacjentów.
- Ograniczone informacje o cechach (zostaną przedstawione na kolejnym slajdzie). Radiografie stomatologiczne i obrazy stomatologiczne wyłącznie ukazują powierzchnię uzębienia wzdłużną, dostarczając informacji o bocznym widoku zębów, lecz brakuje szczegółów dotyczących wewnętrznego układu zębów. Ponadto, większość dostępnych danych niezwiązanych z rentgenem to kolorowe obrazy przednich zębów, które również głównie rejestrują informacje dotyczące ich bocznych aspektów.

# OGRANICZENIA RADIOGRAFII STOMATOLOGICZNYCH



**Fig. 1.** Comparison of the dental X-ray, tooth image, and dental impressions image.

# PRZYKŁADOWA PRÓBKA

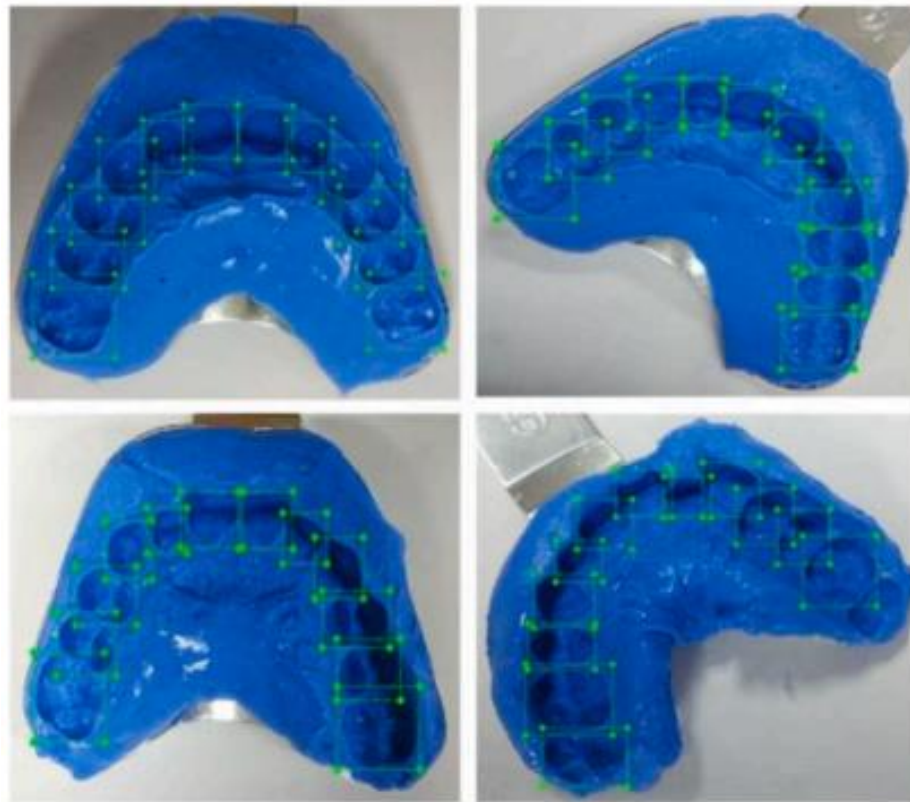
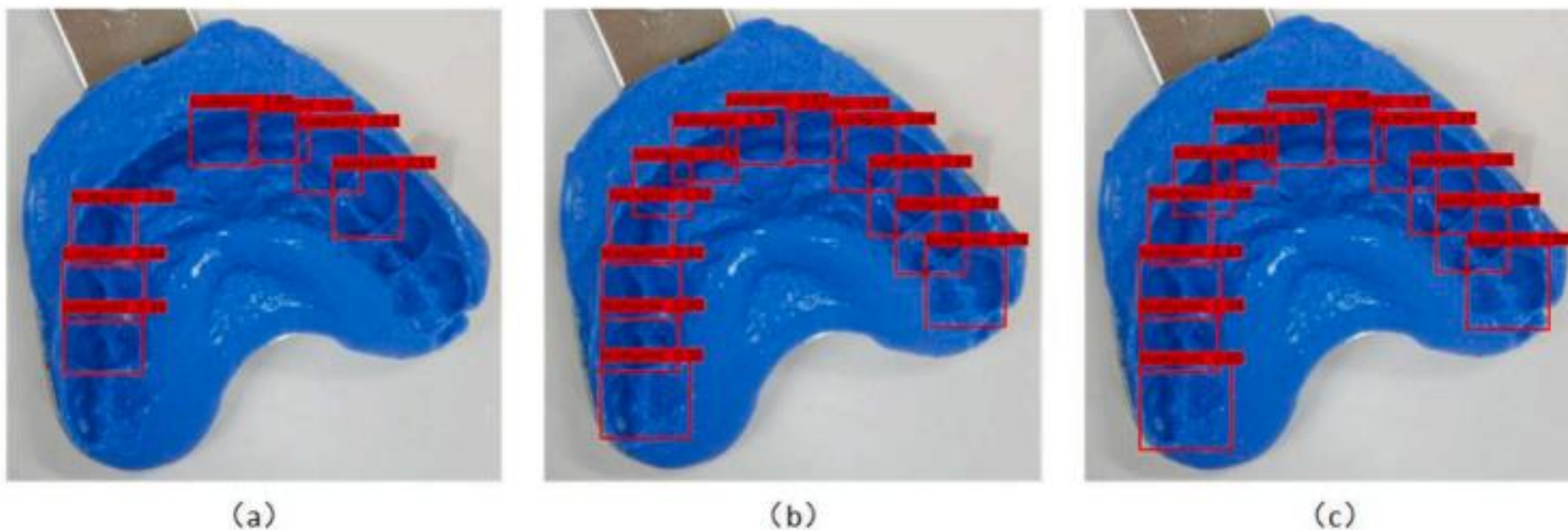


Fig. 2. Labeled sample data.

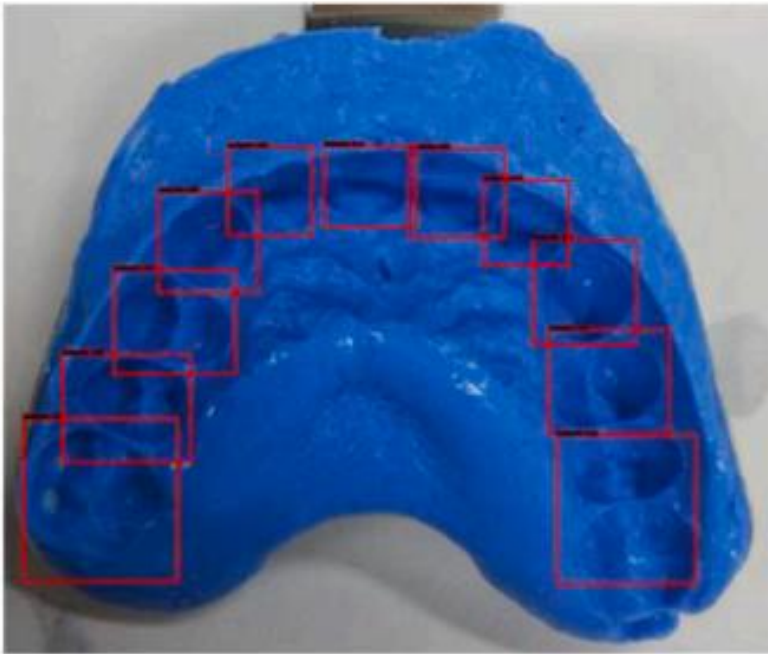
# DOSTOSOWANIE MODELU DO IDENTYFIKACJI NA PODSTAWIE ODCISKÓW ZĘBA



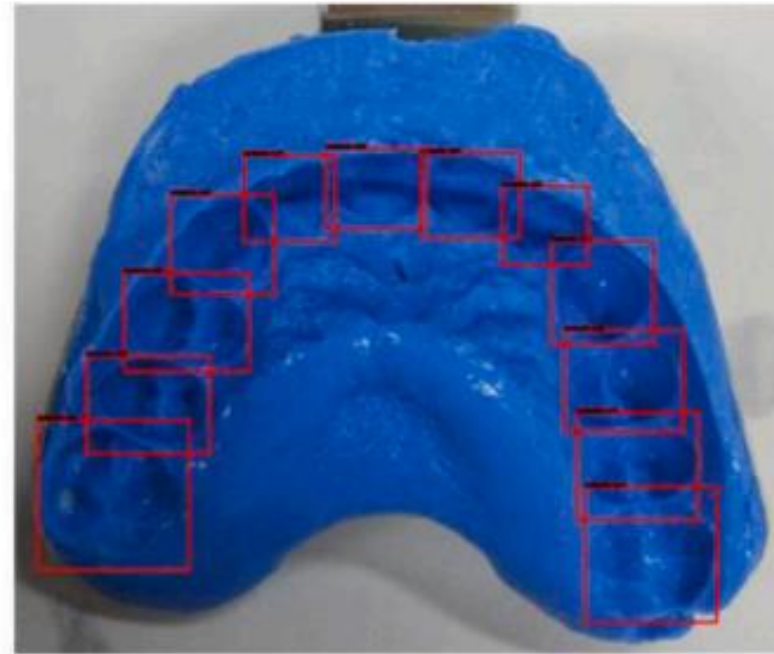
**Fig. 3.** (a) Test result of the 50-person model. (b) Test result of the 30-person model. (c) Test result of the gradient model.



# MODYFIKACJA PODEJŚCIA DODANIEM TZW. PRZEWODNIKA



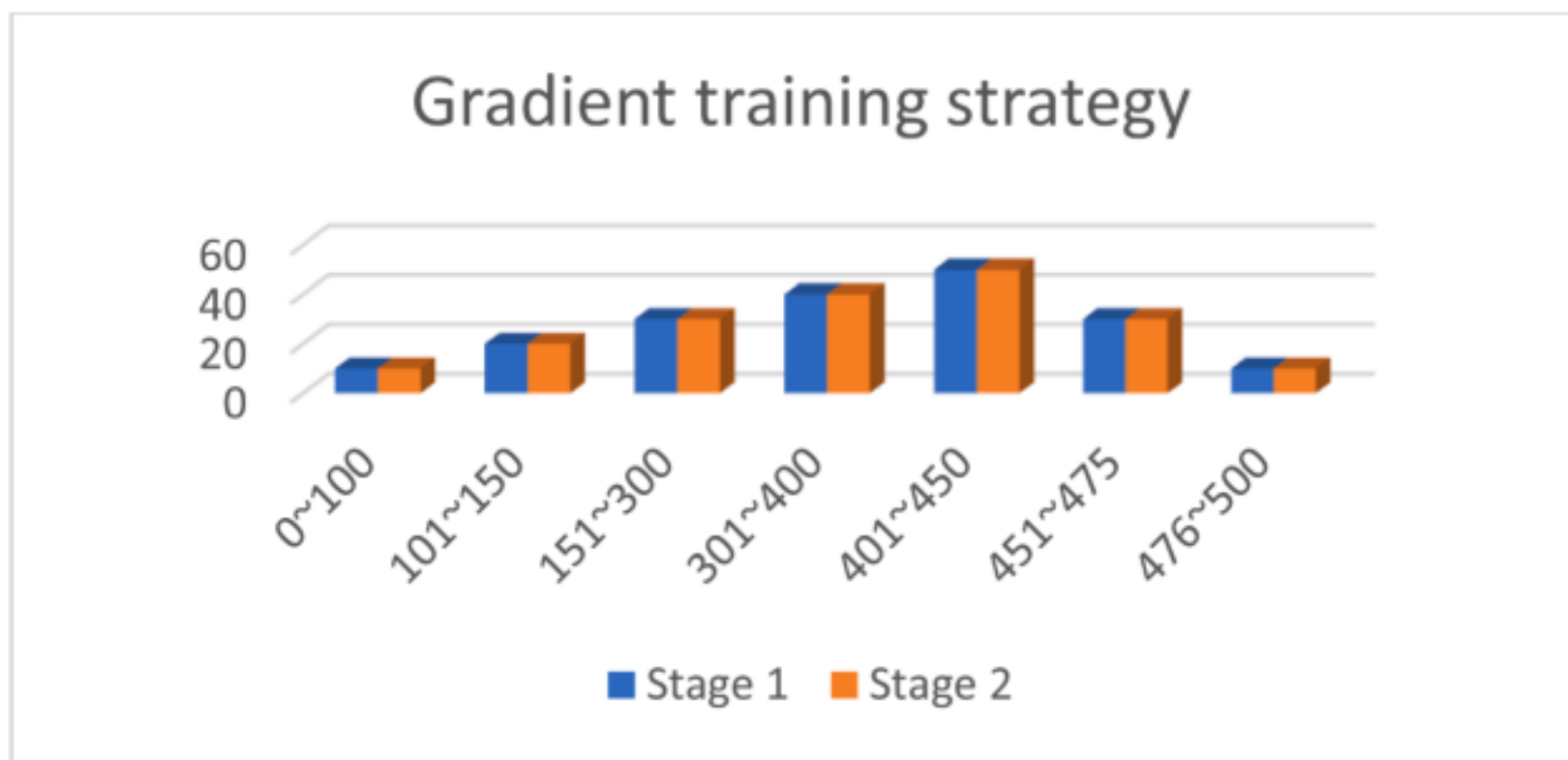
(a)



(b)

**Fig. 4.** (a) The result of detecting No.0301 using 30-person model. (b) The result of detecting No.0301 using gradient model.

# ETAPOWA STRATEGIA GRADIENTOWEGO SZKOLENIA



**Fig. 5.** Gradient training strategy.

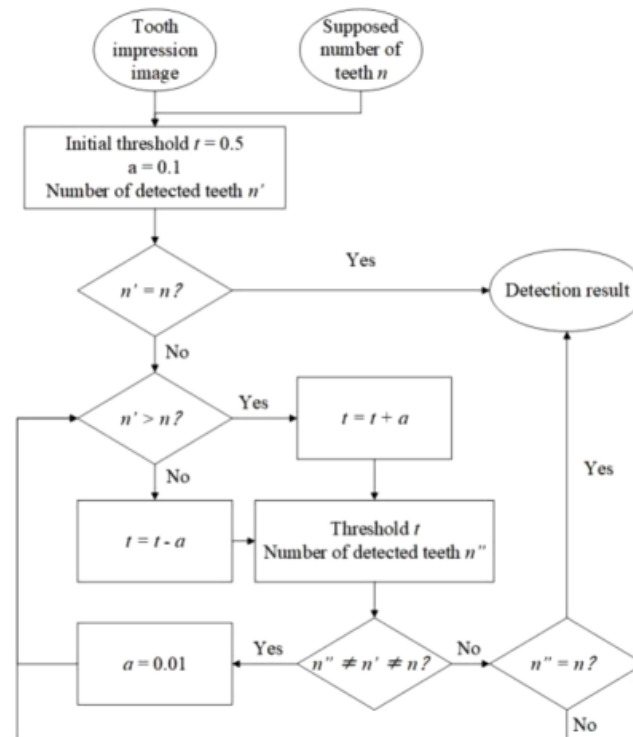


# PROCES ADAPTACYJNEGO PODEJMOWANIA PRÓGU

**Table 1**

Accuracy result of tooth print object detection.

	Fully Detected Images	Total Images	Accuracy
Threshold=0.5	663	916	72.38%
Adaptive threshold	842	916	91.92%



**Fig. 6.** Adaptive threshold decision process.

# WYKRYWANIE "ZŁYCH" ODCISKÓW ZĘBÓW

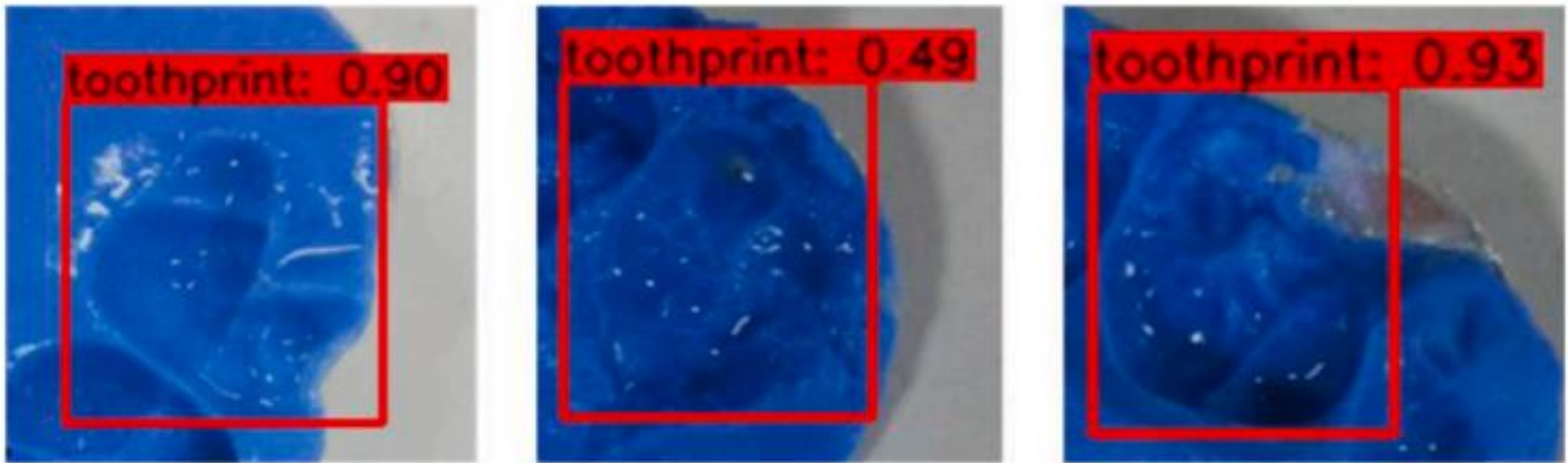
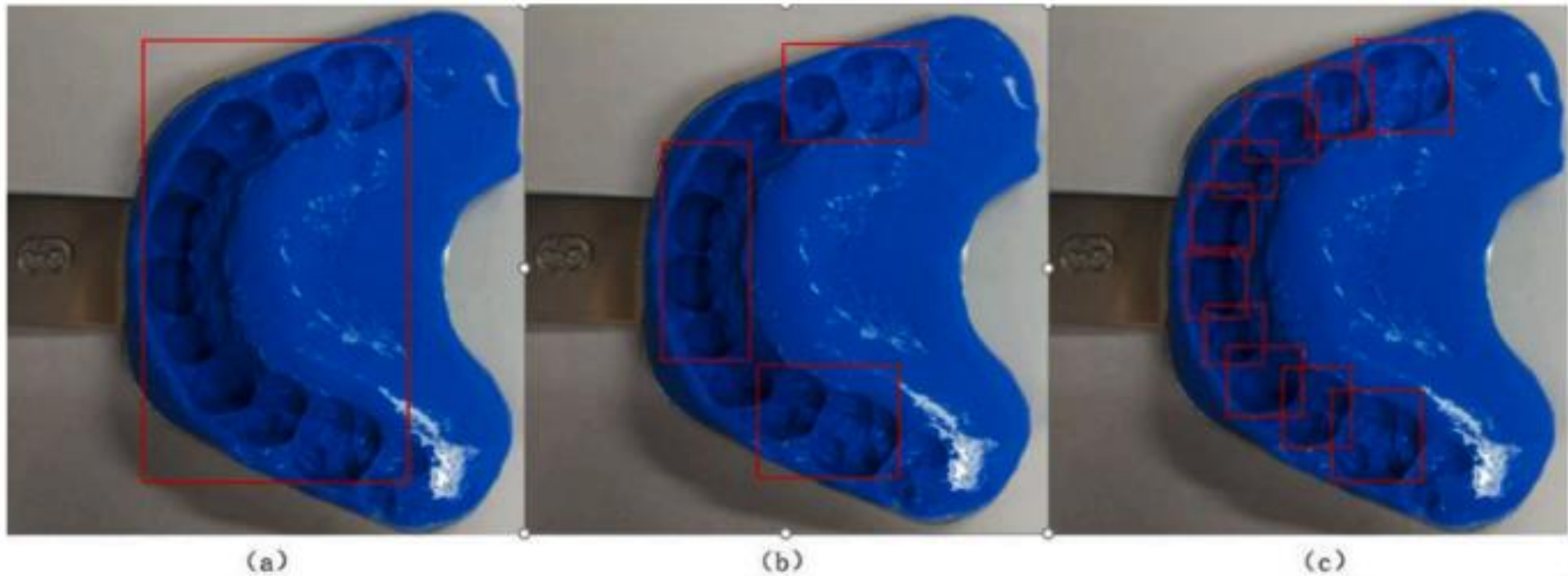


Fig. 7. Detection result of some "bad" tooth prints.

# OBSZARY KLASYFIKACJI ZĘBOWEJ



**Fig. 8.** (a) Global scale area. (b) Local scale areas. (c) Single scale areas.

# PROCES GENEROWANIA LOKALNYCH OBSZARÓW ZĘBOWYCH

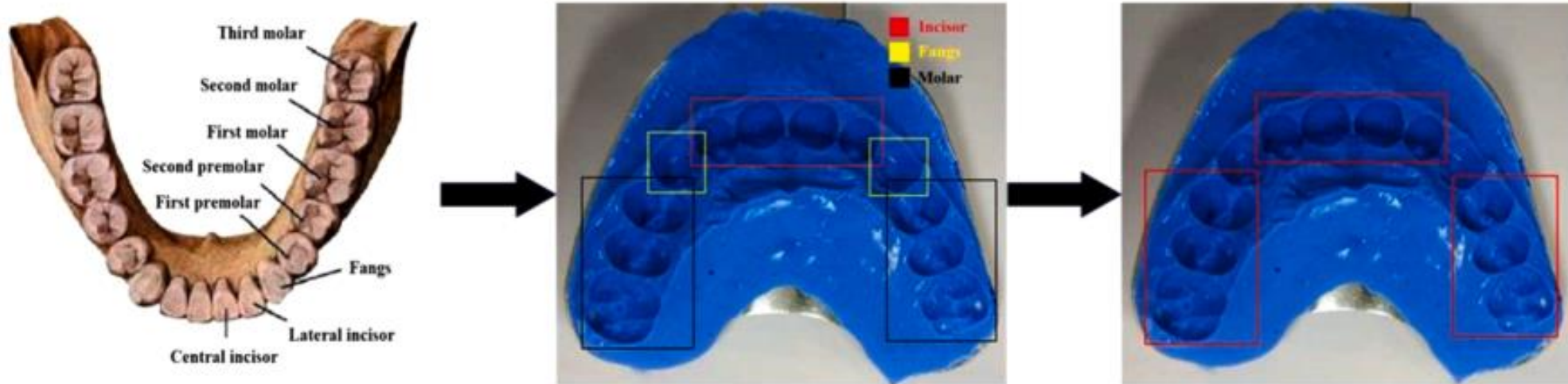


Fig. 9. The process of generating local tooth print areas.

# STRUKTURA ŁUKÓW ZĘBOWYCH

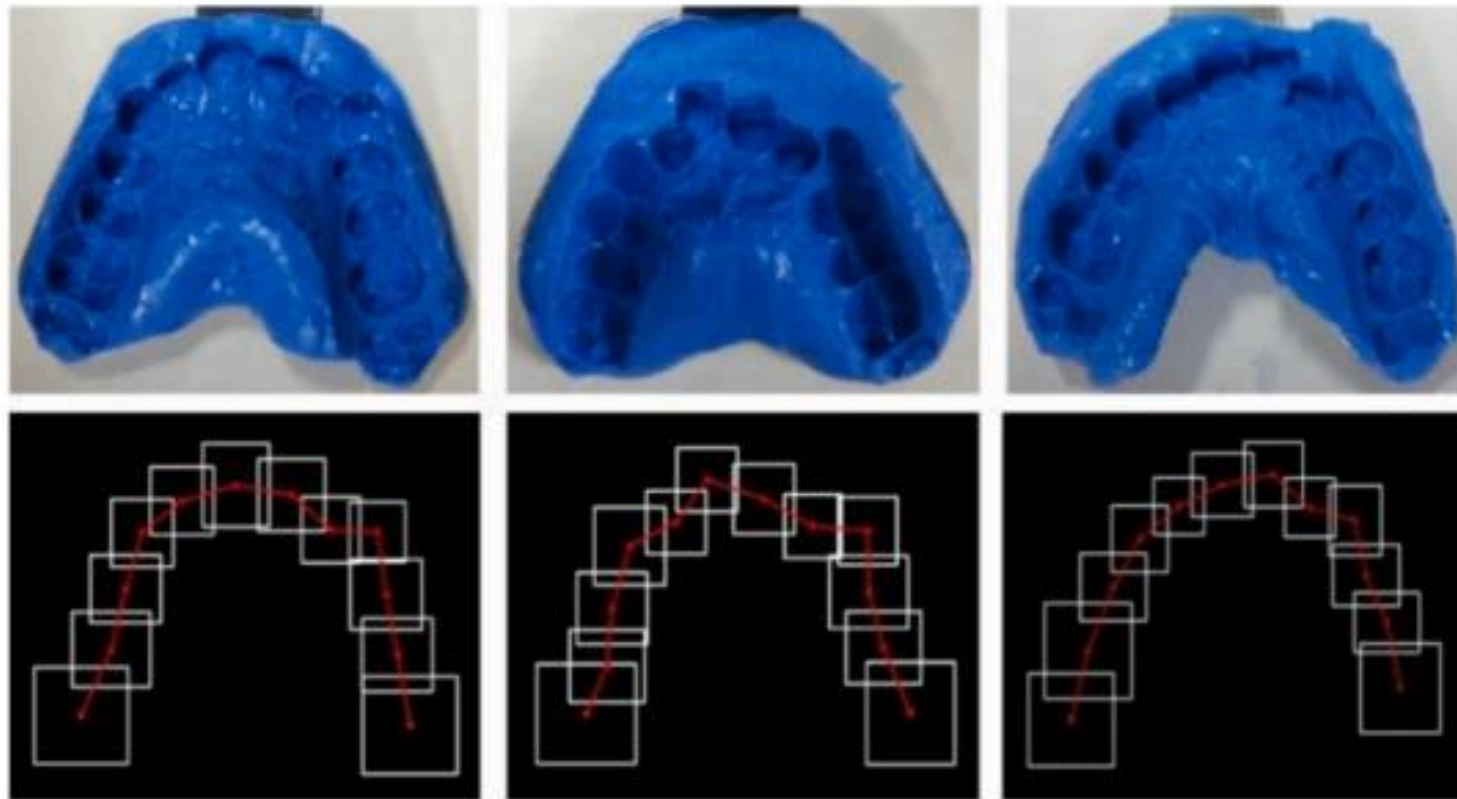
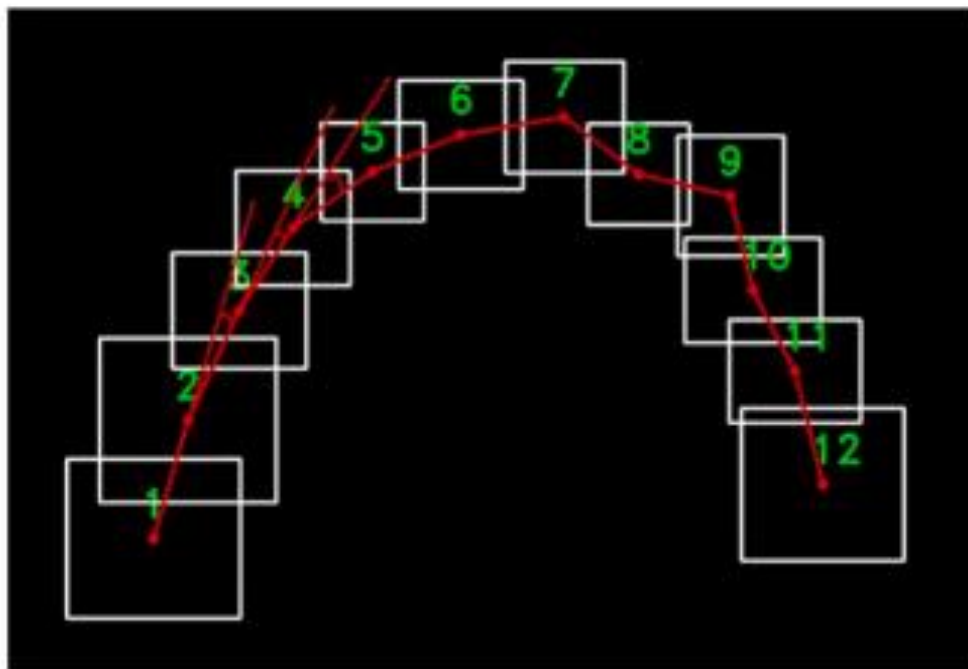
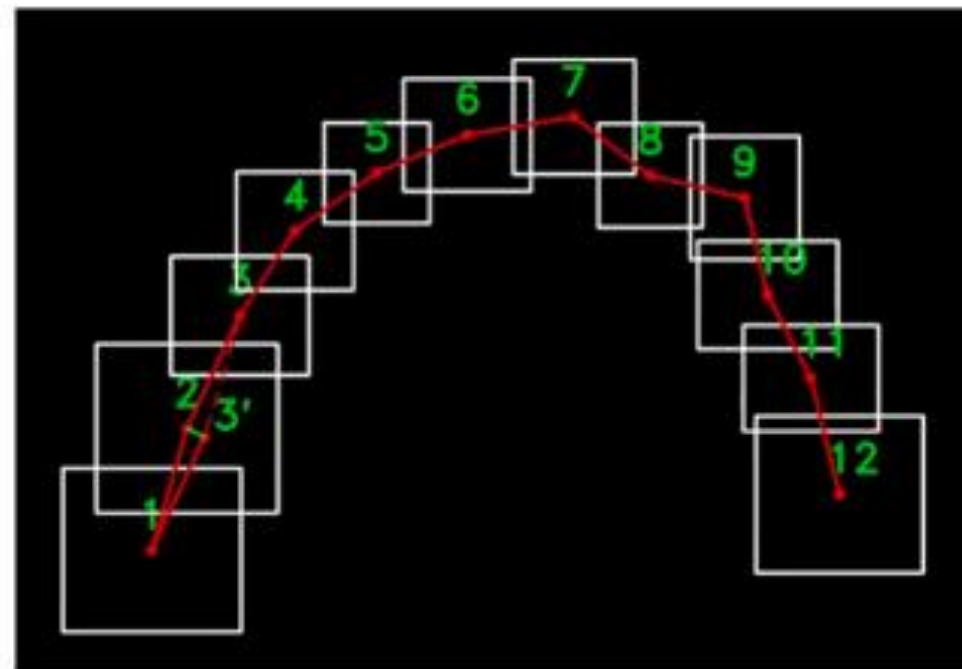


Fig. 10. Overall structures of dental arches.

# OPEROWANIE NA ŁUKACH ZĘBOWYCH



(a)



(b)

**Fig. 11.** (a) Offset angle. (b) Calibrated offset distance.



# RÓŻNICA W ZALEŻNOŚCI OD SEKWENCJI DYSTANSÓW

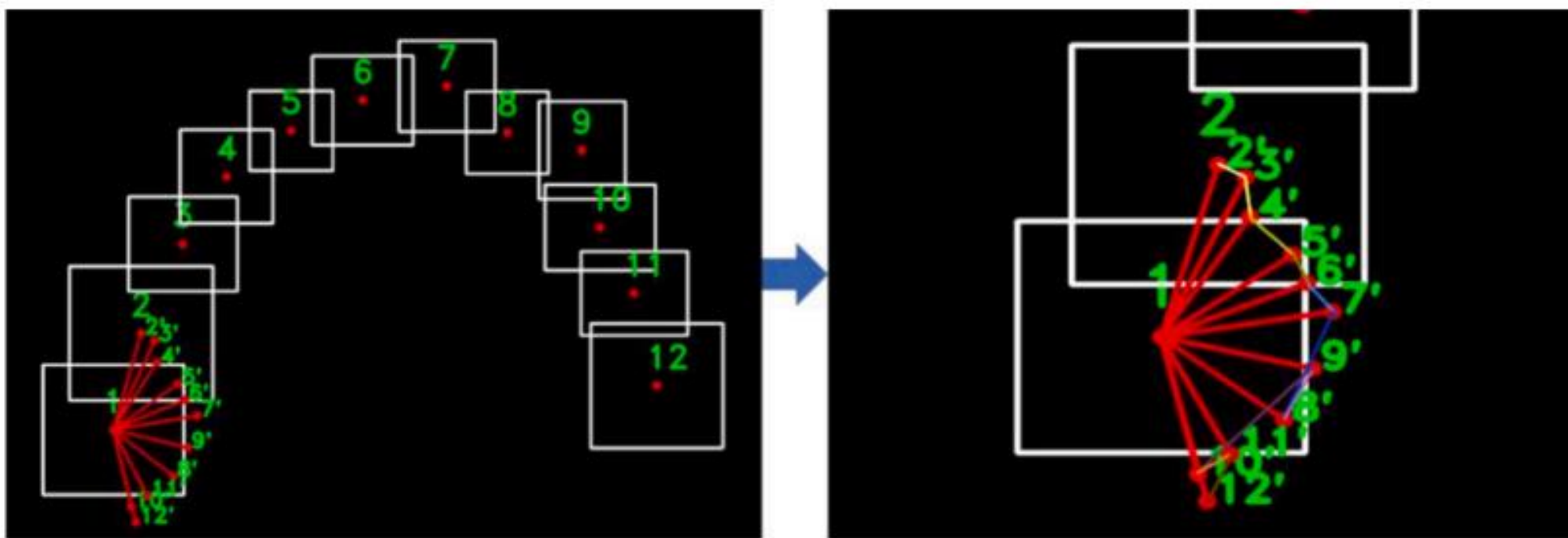


Fig. 12. The calibrated offset distance sequence of a unique individual.

# RÓŻNICE W SZEREGACH SKALIBROWANYCH ODLEGŁOŚCI PRZESUNIĘCIA W ŁUKACH ZĘBOWYCH

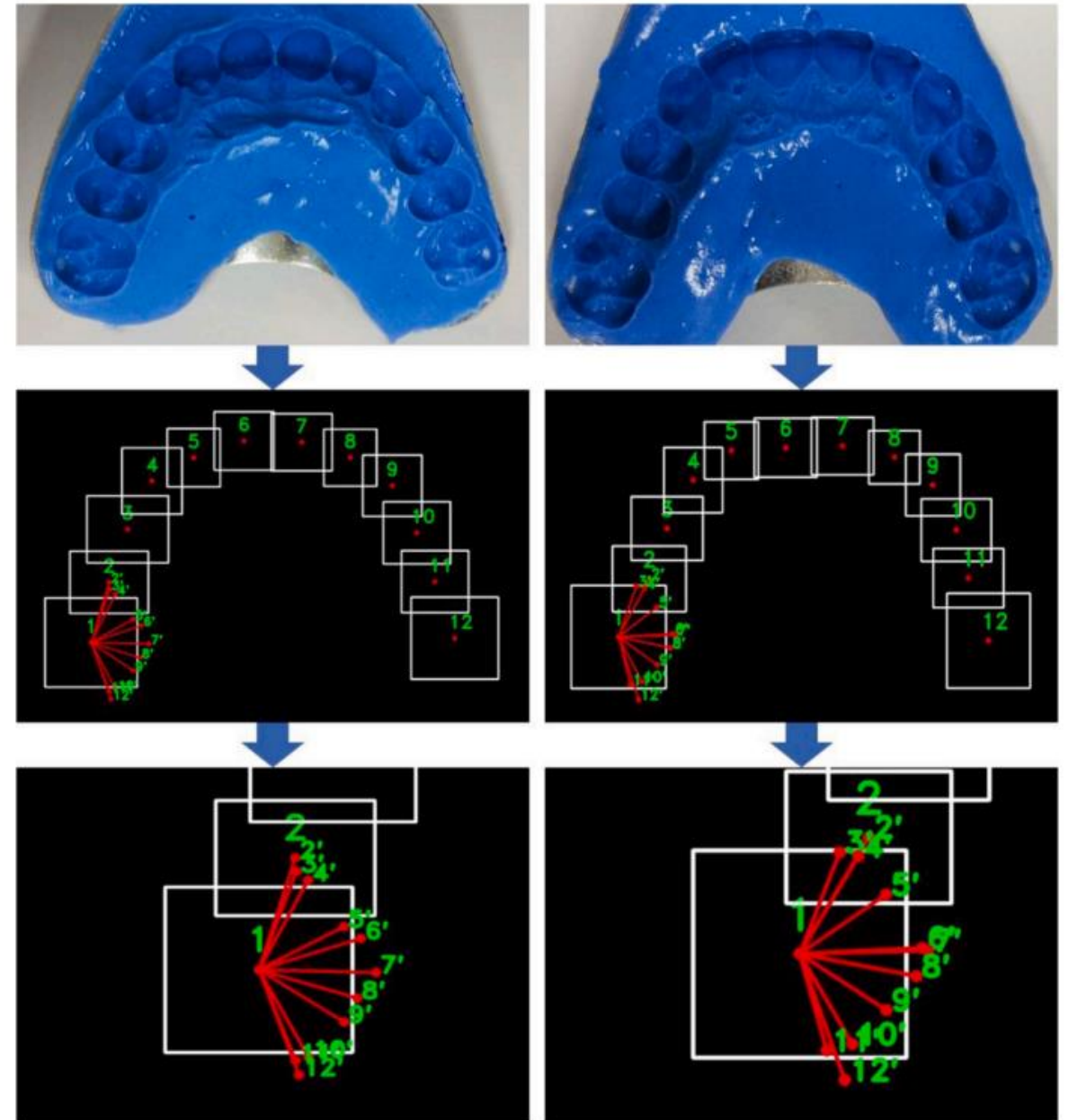


Fig. 13. Calibrated offset distance series of two normal individuals.

# SZCZEGÓŁOWE KROKI OBLICZENIOWE GENEROWANIA SKALIBROWANEJ SERII ODLEGŁOŚCI OFFSETOWEJ

- (1) Obtain  $n$  tooth print bounding boxes on a single scale of an input dental impression image, and assign numbers to the center points of these boxes as  $1 \sim n$ . The coordinates of each point  $N_n$  are as follows:

$$N_n = (x_n, y_n)$$

- (3) It can be calculated the calibrated offset distances  $\dot{d}_n$  :

$$\dot{d}_n = \sqrt{(\dot{x}_n - \dot{x}_{n-1})^2 + (\dot{y}_n - \dot{y}_{n-1})^2}, n \geq 3$$

Calibrate the offset distance series  $\dot{d}_k (k = 3, 4, \dots, n)$  to form a one-dimensional vector, denoted as  $\vec{d}$ . The distance vector can be obtained as follows:

$$\vec{d} = [\dot{d}_3, \dot{d}_4, \dot{d}_5, \dots, \dot{d}_n], n \geq 3$$

- (2) Taking point No. 1 ( $x_1, y_1$ ) as the base point, translate each fold line segment to obtain a series of projections, including  $n - 1$  points. The coordinates of each projected point  $N'_n$  are as follows:

$$N'_n = (\dot{x}_n, \dot{y}_n) = (x_n - (x_{n-1} - x_1), y_n - (y_{n-1} - y_1)), n \geq 2$$

- (4) Divide  $(\vec{d} - \dot{d}_3)$  by  $\dot{d}_3$  to obtain the scaling factor for calibration. This results in the calibrated offset distance vector, denoted as  $\overrightarrow{bias\_d}$ :

$$\overrightarrow{bias\_d} = \left[ \frac{\dot{d}_4}{\dot{d}_3}, \frac{\dot{d}_5}{\dot{d}_3}, \dots, \frac{\dot{d}_n}{\dot{d}_3} \right], n \geq 4$$

# EKSPANSJA W ZALEŻNOŚCI OD ILOŚCI ODCISKÓW ZĘBÓW

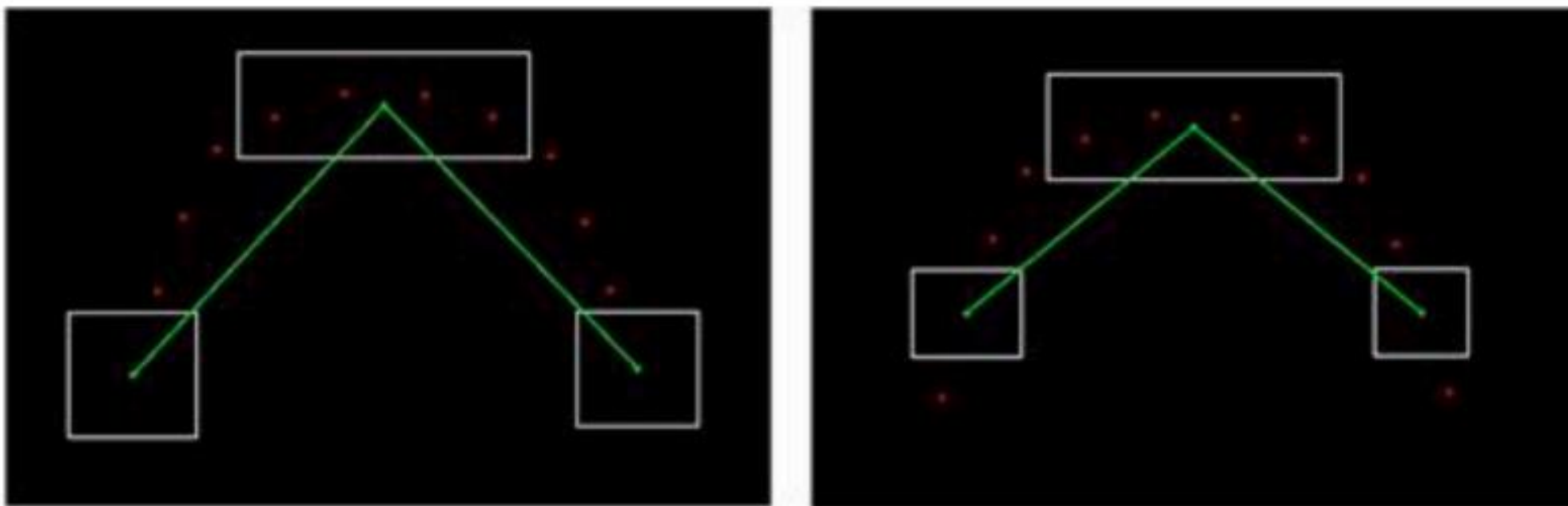


Fig. 14. (a) Expansion angle of 12 tooth prints. (b) Expansion angle of 10 tooth prints.

# PORÓWNANIE PUNKTÓW CECHOWYCH SIFT

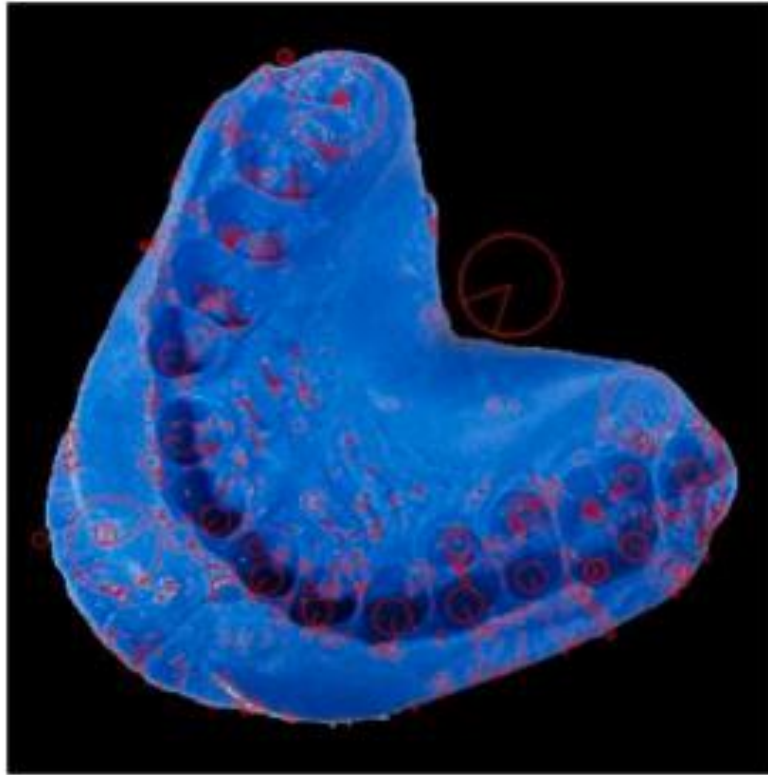
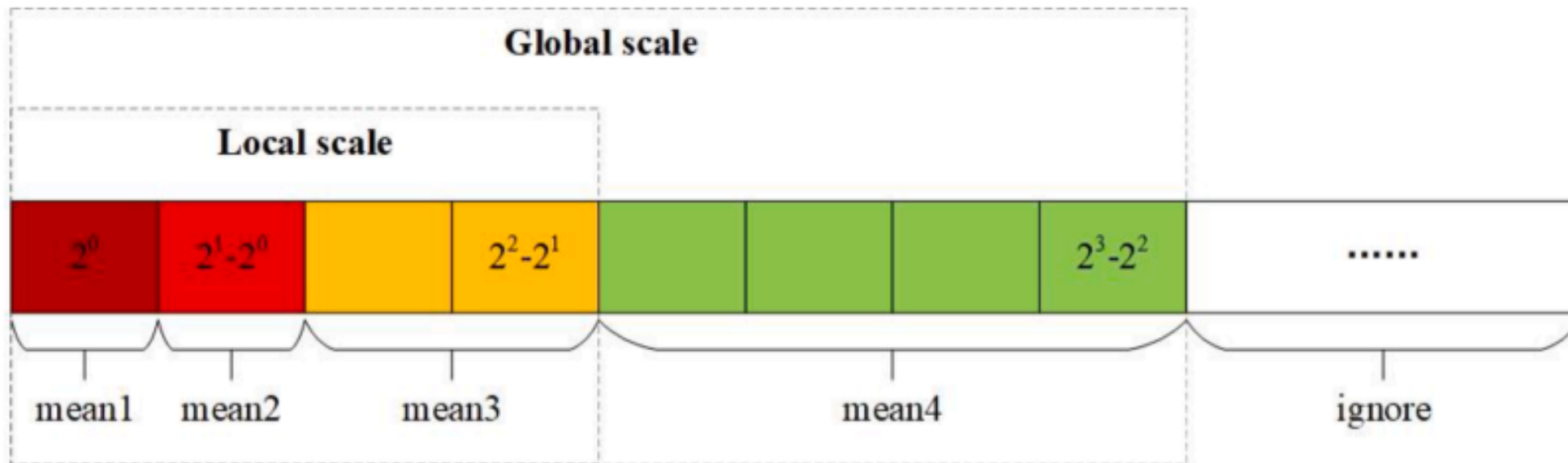


Fig. 15. The SIFT feature points of the dental impression image.

# SELEKCJA I OBLICZANIE CECH PODOBIEŃSTWA OPARTYCH NA SIFT



**Fig. 16.** Multi-scale tooth print features based on SIFT.



# PODZIAŁ DATASETU

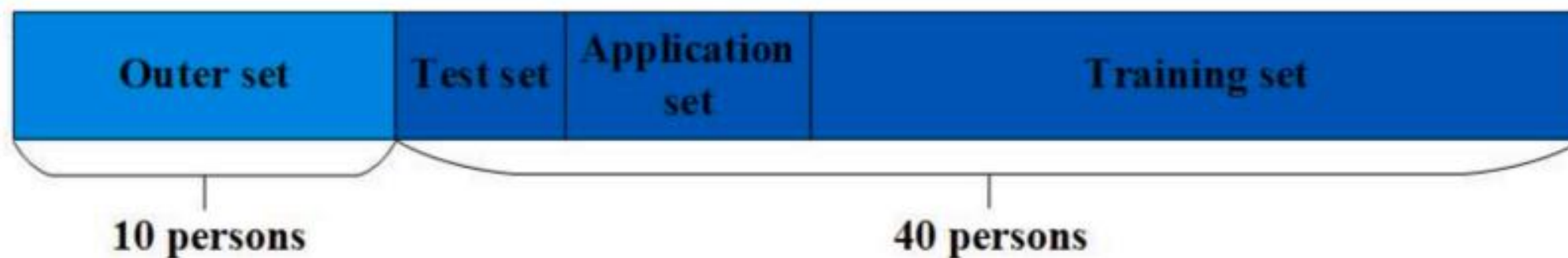


Fig. 17. The structure of the data set.

# ZMIANA KLASYFIKACJI W IDENTYFIKACJI WYCISKÓW DENTYSTYCZNYCH

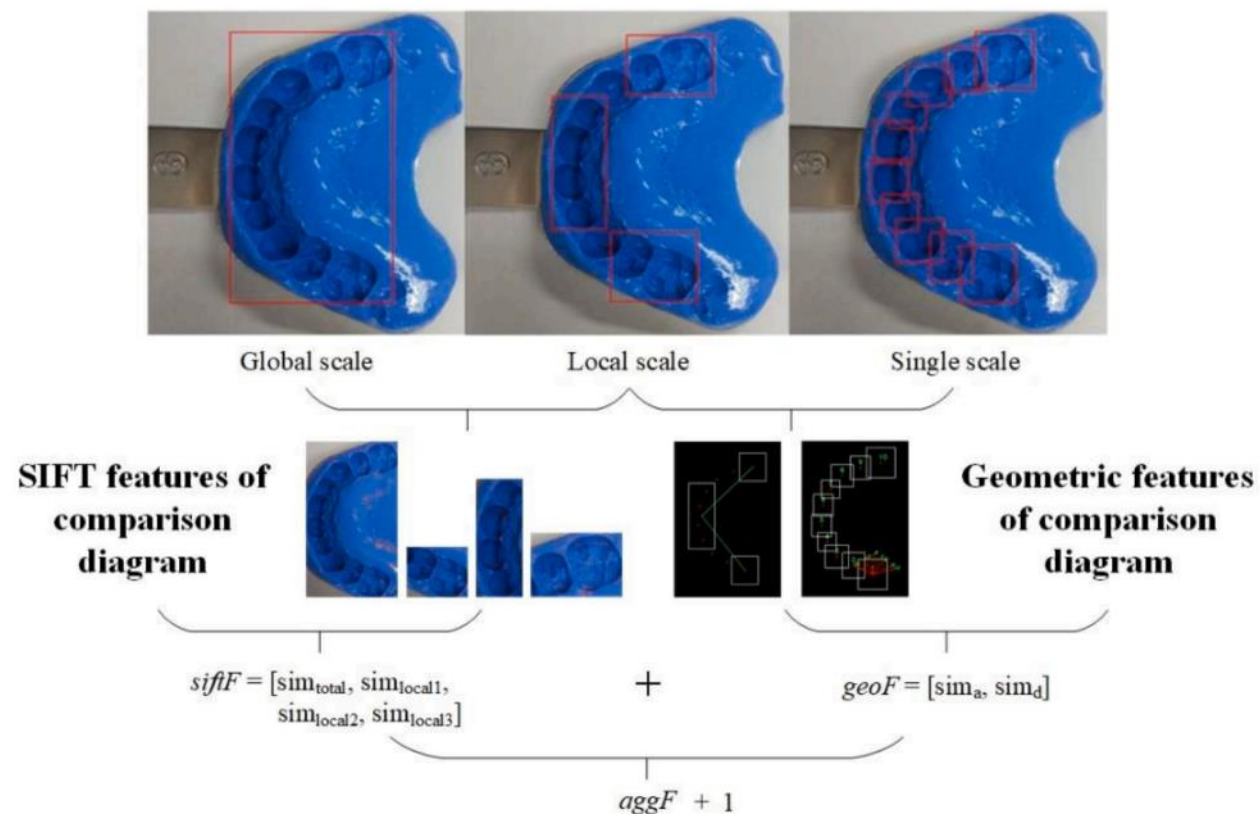


Fig. 18. Construction process of a positive sample.

# MODEL TESTOWY



Próbka "pozytywna" ("1"): dwa obrazy należą do tej samej osoby



Próbka "negatywna" ("0"): dwa obrazy należą do różnych osób



Przez dokładność testu rozumiemy poprawność identyfikacji próbek pozytywnych i negatywnych i pomiar dokładności ich klasyfikacji.

# ZBIÓR DANYCH TESTOWYCH



40 osób. Po dwa zdjęcia dla każdej z osób, jedno pozytywne, drugie negatywne.



Co daje finalnie **80 zdjęć**.

# WYNIKI TESTÓW

Table 4. Test results and comparison.

	Geometry features	Aggregation features
Test Accuracy	72/80 90.00%	76/80 95.00%
Positive sample accuracy	38/40 95.00%	38/40 95.00%
Negative sample accuracy	34/40 85.00%	38/40 95.00%

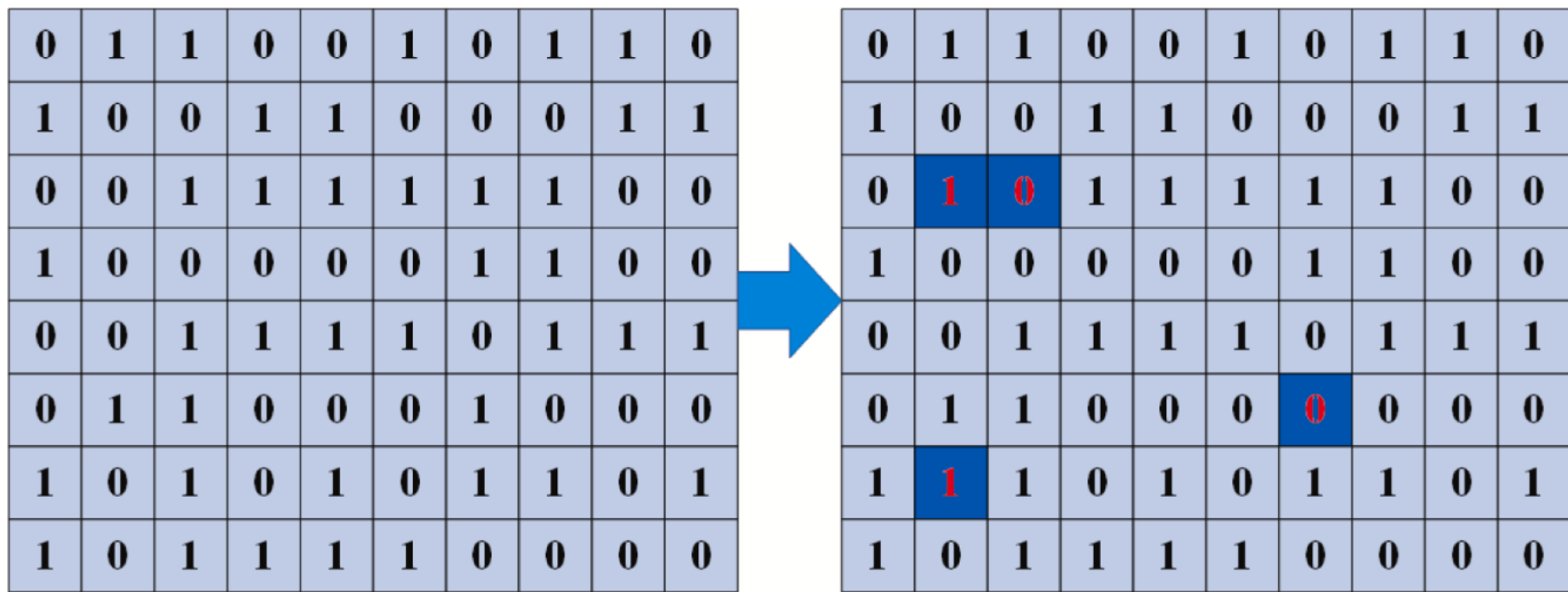


Fig. 19. Test sample labels.



# MODEL W ZASTOSOWANIU PRAKTYCZNYM

Autorzy porównali swoją metodę identyfikacji opartą na zdjęciach szczęki, z pokrewnymi metodami innych naukowców w innych pracach.

TOP-N (dla N-naturalnego) oznacza "poziom ufności", im niższa wartość N, tym bardziej restrykcyjne kryteria rozpoznawania uznajemy.

W omawianej pracy, do eksperymentu praktycznego zastosowano zbiór 40 zdjęć.

# WYNIKI

**Table 5**  
Comparison of dental identification accuracy.

Method	Jaw	TOP-N retrievals		
		<i>N</i> = 1	<i>N</i> = 3	<i>N</i> = 5
[44]	upper	130/311 41.80%	229/311 73.63%	268/311 86.17%
[48]	upper	50/115 43.50%	77/115 66.90%	88/115 76.50%
	lower	50/105 47.60%	72/105 68.60%	83/105 79.00%
	both	87.21%	—	95.34%
Ours	upper	34/40 85.00%	40/40 100%	40/40 100%



Scenariusz 1: Porównanie i identyfikacja dwóch zdjęć dentystycznych, które nie zostały wprowadzone do bazy danych, w celu ustalenia, czy należą one do tej samej osoby. Oczekiwany wynik to „1” lub „0”, co wskazuje, czy pochodzą one od tej samej osoby, czy nie.



Scenariusz 2: Identyfikacja obrazu dentystycznego, którego prawdziwej tożsamości nie ma w bazie danych. Oczekiwany wynik to NONE, co oznacza, że w bazie danych nie ma odpowiedniej tożsamości dla danego obrazu. Klasyfikujemy wynik wykrywania jako „brak”, gdy prawdopodobieństwo jest mniejsze niż 0.5.

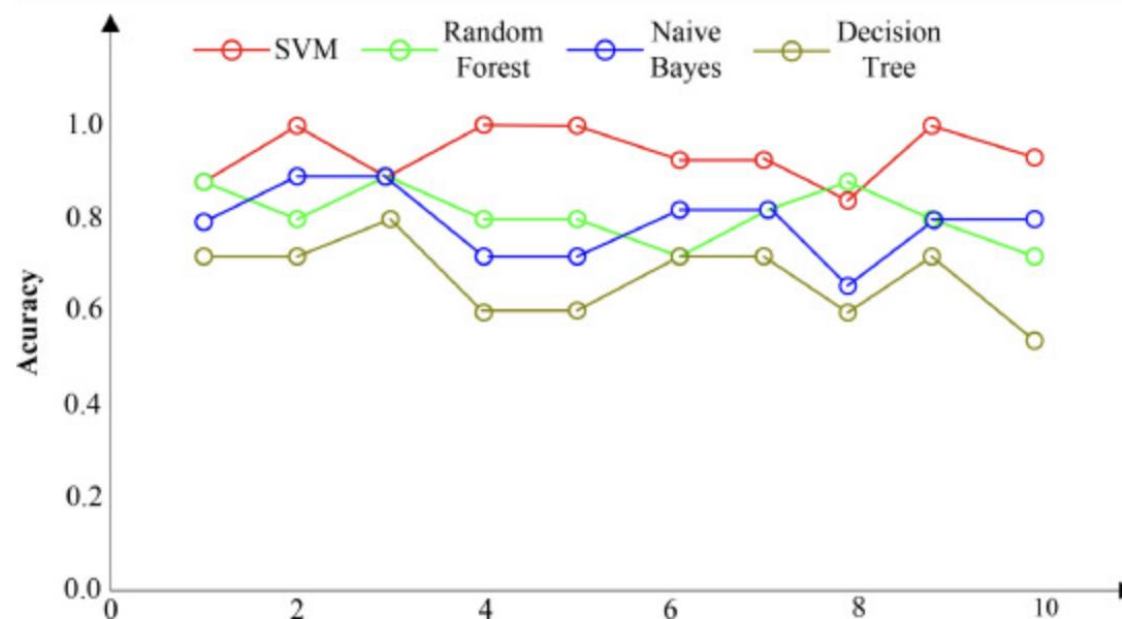
## INNE EKSPERYMENTY

# WYNIKI DLA DWÓCH SCENARIUSZY

	Number of samples	Accuracy
Scenario 1	1487/1764	84.30%
Scenario 2	61/100	61.00%

# POSŁOWIE: PORÓWNANIE POPULARNYCH METOD KLASYFIKACJI W KONTEKŚCIE IDENTYFIKACJI OBRAZÓW ZĘBOWYCH

Autorzy postanowili przeprowadzić jeszcze analizę porównawczą efektywności czterech popularnych klasyfikatorów: SVM, Random Forest, Naive Bayes, Decision Tree.



# PODSUMOWANIE



Autorzy przedstawili motywację i metody identyfikacji osób opartą na zdjęciach stomatologicznych



W artykule wykorzystano sieć YOLOv8, zaprojektowano strategię uczenia gradientowego i zaproponowano strategię adaptacyjnego progu, co dało współczynnik dokładności na poziomie 91,92%



Funkcje zaproponowane w tym artykule osiągnęły dokładność walidacji na poziomie 93,38% w zbiorze danych szkoleniowych, dokładność na poziomie 93,25% w zbiorze danych testowych i współczynnik rozpoznawania na poziomie 100% w przykładowej aplikacji



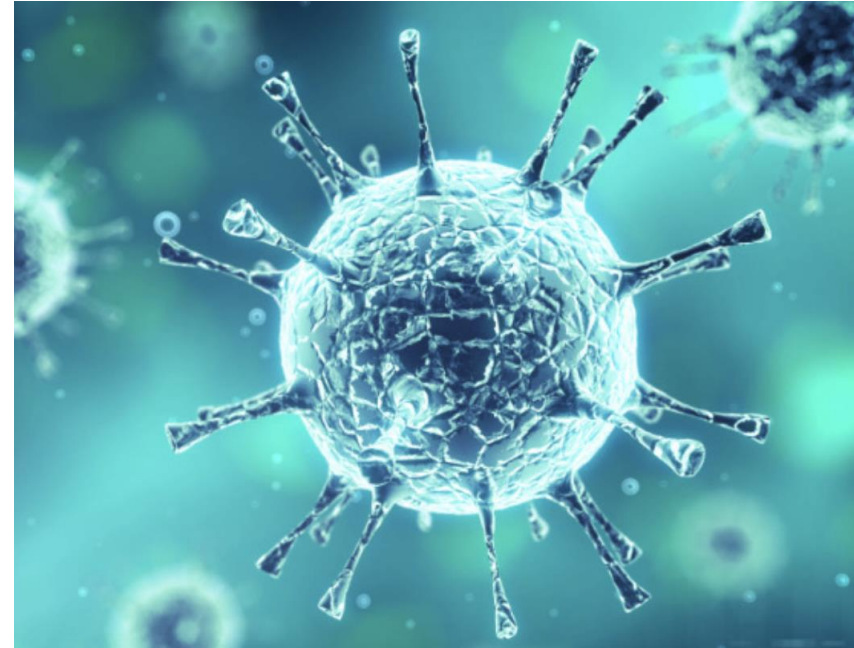
# PRZYSZŁE BADANIA

Wyodrębnione cechy nie są wystarczająco kompleksowe, ponieważ nie oddają ważnych cech zębów, takich jak głębokość i powierzchnia otworu zębowego. Zwrócenie uwagi na te parametry, pomogłoby jeszcze bardziej wzmocnić jakość identyfikacji.

Warto zastanowić się też nad analizą całościową powierzchni zębowej (rozpoznawanie cech szczególnych itp.).

Do przyszłych badań wymaga się o wiele większej bazy danych obrazów. Niedostateczna liczba danych to jeden z głównych problemów badań w tym zakresie.

# PRACA NR 2



## Momentum contrast transformer for COVID-19 diagnosis with knowledge distillation

Aimei Dong<sup>1</sup>, Jian Liu<sup>1</sup>, Guodong Zhang, Zhonghe Wei, Yi Zhai, Guohua Lv\*

*Faculty of Computer Science and Technology, Qilu University of Technology (Shandong Academy of Sciences), Jinan 250353, China*

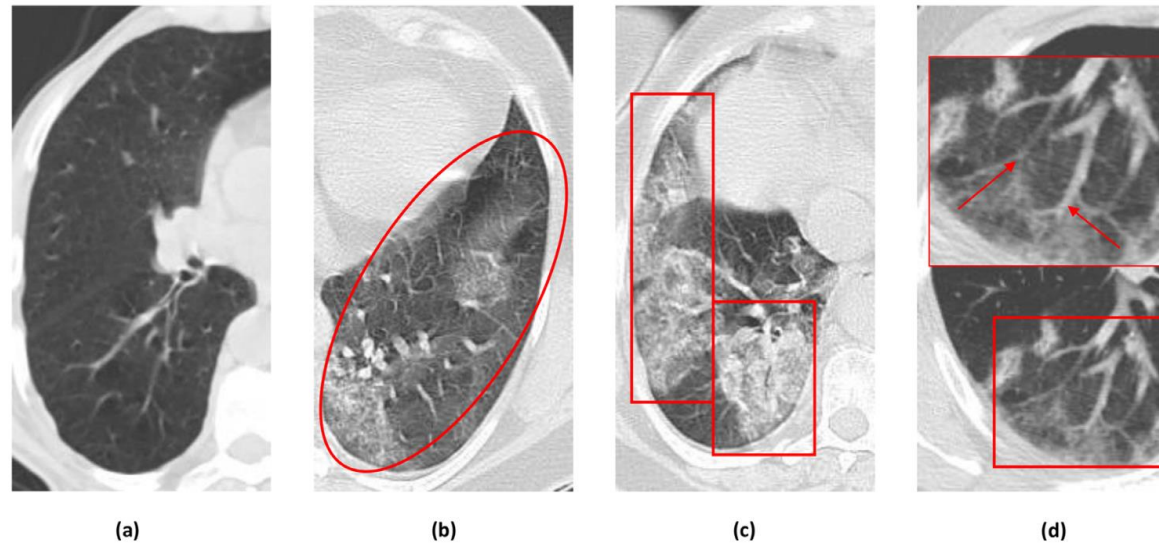
# ABSTRACT — CZEGO SIĘ DOWIEMY?

Autorzy wprowadzają nowatorski model zwany MCT-KD służący diagnostyce COVID-19 w oparciu o zdjęcia tomograficzne klatki piersiowej.

Wykorzystywanie wielu cech szczególnych sprawia, że proponowany model osiąga znakomite wyniki w diagnozie tej choroby, pomimo względnie małej liczbie danych do uczenia i testów.

# WSTĘP

COVID-19 jest chorobą zakaźną i stanowi duże zagrożenie dla życia ludzkiego. Zdjęcia tomografii komputerowej w zakresie klatki piersiowej są bezinwazyjne i odgrywają ważną rolę w diagnostyce choroby.



**Fig. 1.** Comparison of chest CT images between Healthy Control (a) and COVID-19 patients (b)-(d). (b) Large area of ground glass opacities (red circle). (c) Crazy-paving pattern (red rectangle). (d) Local bronchiolectasis (red arrow).

# DOTYCHCZASOWE METODY DIAGNOSTYKI COVID-19 NA PODSTAWIE OBRAZÓW MEDYCZNYCH

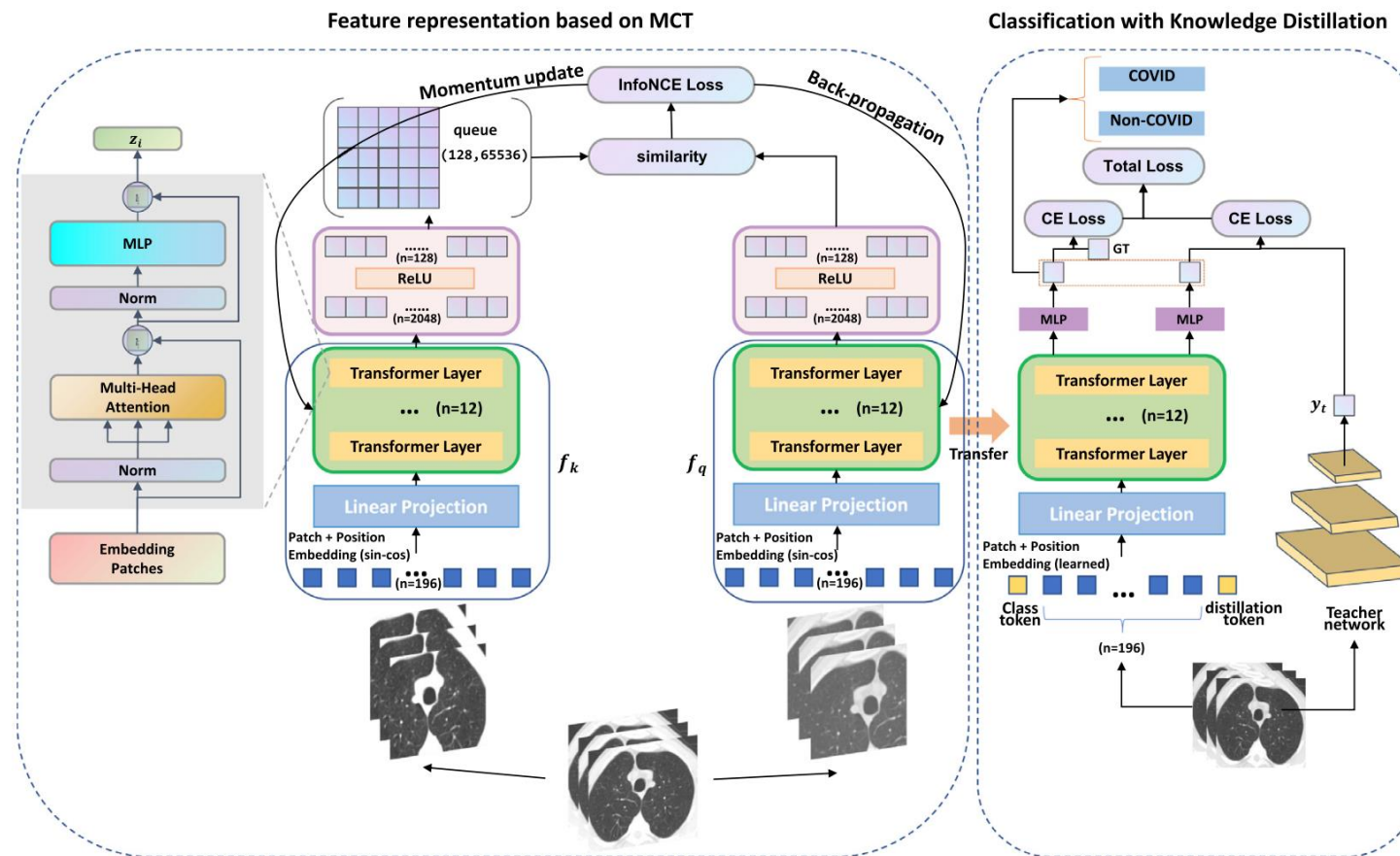
Główne idee uczenia maszynowego stosowane do diagnostyki COVID-19 możemy podzielić na:

- Klasyfikację opartą na CNN (Convolutional neural network)
- Klasyfikację opartą na transformerach

Propozycja reprezentacji cech opartej na zadaniu MCT, w celu wydobywania globalnych cech. To zadanie to uczenie samo-nadzorowane, które może rozwiązać problem trudności w szkoleniu ViT na małych zestawach danych. Drugi etap to projekt klasyfikacji z destylacją wiedzy, aby wprowadzić założenie lokalności CNN - dzięki czemu możemy skupić się zarówno na lokalnych, jak i na globalnych cechach.

## WSTĘP DO IDEI

# STRUKTURA MCT-KD





# AUGMENTACJA DANYCH UŻYTA W PROJEKCIE

Autorzy używają losowych metod augmentacji danych, mianowicie: losowego wycinania, losowego drżenia kolorów, losowej skali szarości, losowego rozmycia gaussowskiego, losowego poziomego odbicia. Przykład augmentacji danych jest pokazany poniżej:

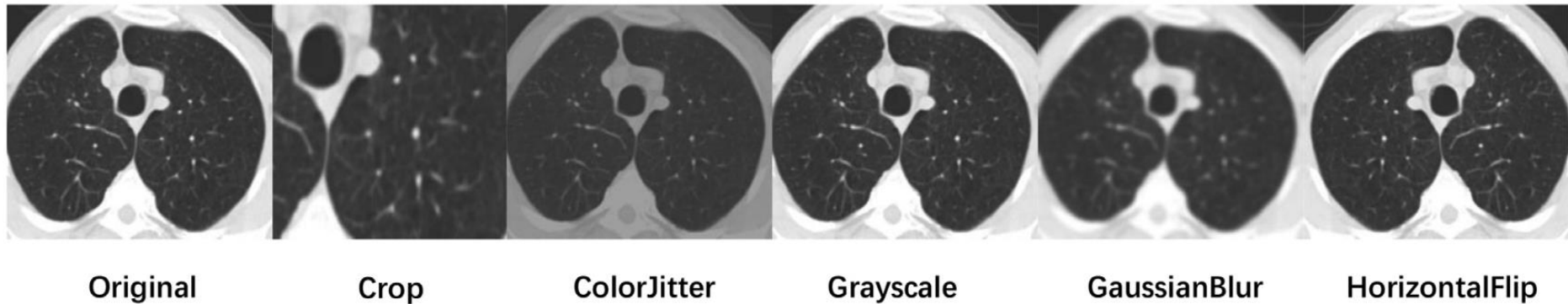
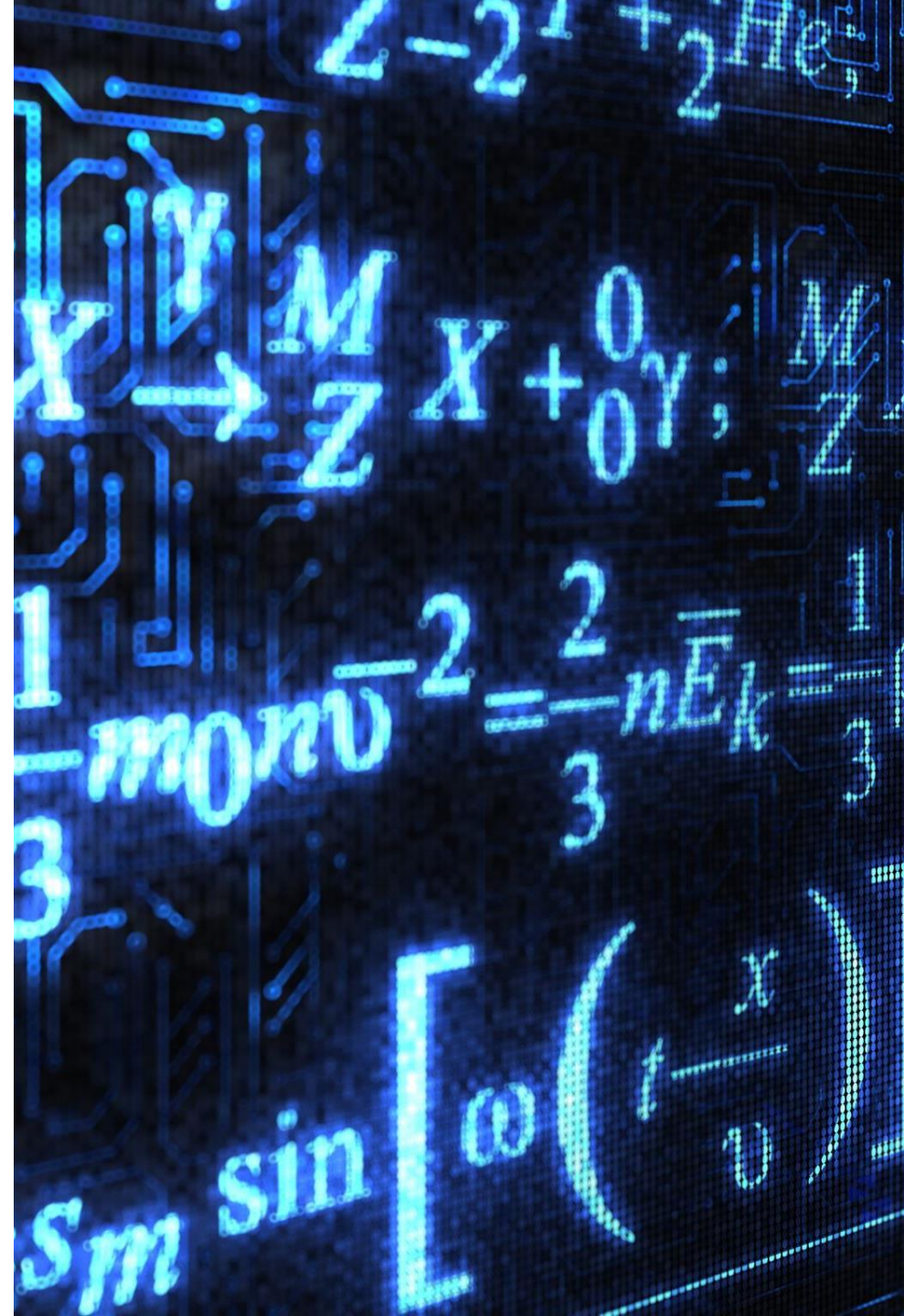


Fig. 3. Five methods of data augmentation.

# KONTRASTOWA NAUKA REPREZENTACJI

Kontrastowa nauka reprezentacji składa się z czterech modułów:

- Ekstrakcja cech (feature extraction)
- Głowica projekcyjna (projection head)
- Kontrastowa funkcja straty (contrastive loss function)
- Kontrast pędu (momentum contrast)



# EKSTRAKCJA CECH

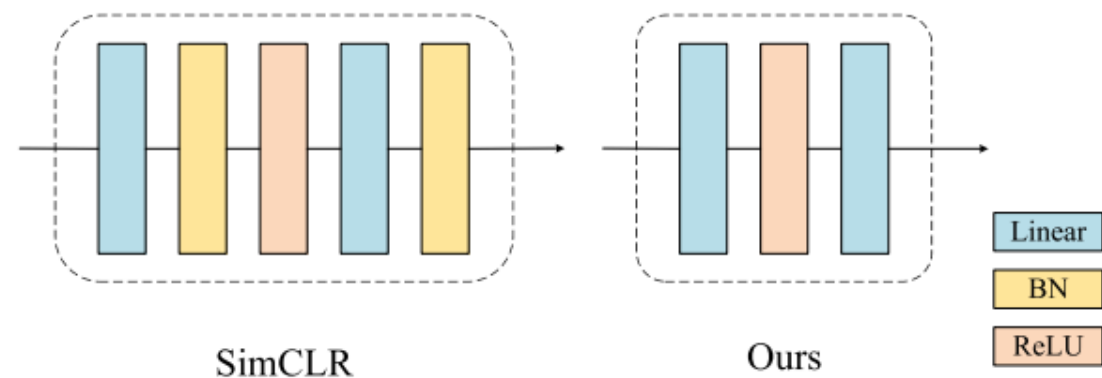
Ze względu na ograniczone publiczne zbiory danych COVID-19, ViT został wprowadzony do uczenia kontrastywnego z momentem, aby uzyskać ukryte cechy. Konkretnie, ViT służy jako enkoder do analizy obrazów CT klatki piersiowej COVID-19.

# GŁOWICA PROJEKCYJNA

Głowica projekcyjna mapuje reprezentację z ekstrakcji cech do przestrzeni cech (feature space) do obliczenia kontrastowej straty.

W porównaniu do SimCLR głowica projekcyjna nie korzysta z warstwy BN, gdyż zawiera tylko dwie warstwy liniowe i jedną ReLU. Według badań, warstwa BN może prowadzić do wycieku informacji, przez co model zaczął uczyć się skrótów.

Uzyskana cecha służy do obliczenia kontrastowej straty.



**Fig. 4.** Comparison between SimCLR and our projection head.



Podstawą uczenia kontrastycznego jest zadanie wyszukiwania w słowniku. Dla  $N$  losowo zakodowanych próbek w mini-batchu, uzyskuje się  $2N$  próbek z wyniku głowicy projekcyjnej. Dla losowej próbki  $z_q$  reszta to "słownik".



Celem uczenia kontrastycznego jest zakodowanie próbki  $z_q$ , by znaleźć jej pozytywną próbkę w słowniku. Ta sama próbka to próbka pozytywna, w przeciwnym wypadku negatywna.

## KONTRASTOWA FUNKCJA STRATY

Uczenie kontrastywne wymaga bardzo dużej liczby próbek negatywnych. Tradycyjne metody używają mini-batcha jako słownika próbek, co z kolei wymaga dużo zasobów sprzętowych.



Problem ten jest rozwiązalny przez zastosowanie kontrastu z momentem. Jego istota polega na utrzymaniu kolejki jako słownika próbek. Dzięki właściwości First In, First Out (FIFO), klucze z poprzednich batchy mogą być ponownie wykorzystywane. Gdy słownik jest pełny, stare rekordy są usuwane, a najnowsze zachowywane. Liczba próbek negatywnych zależy więc nie od mini-batcha, ale od rozmiaru słownika. Dzięki temu używamy standardowego rozmiaru batcha, ale słownik może być bardzo duży.

## KONTRAST PĘDU



Aby poprawić wydajność metody i uwzględnić lokalne cechy COVID-19, jest stosowana destylacja wiedzy (knowledge distillation) podczas klasyfikacji. CNN są znacznie lepsze do ekstrakcji cech lokalnych niż ViT.



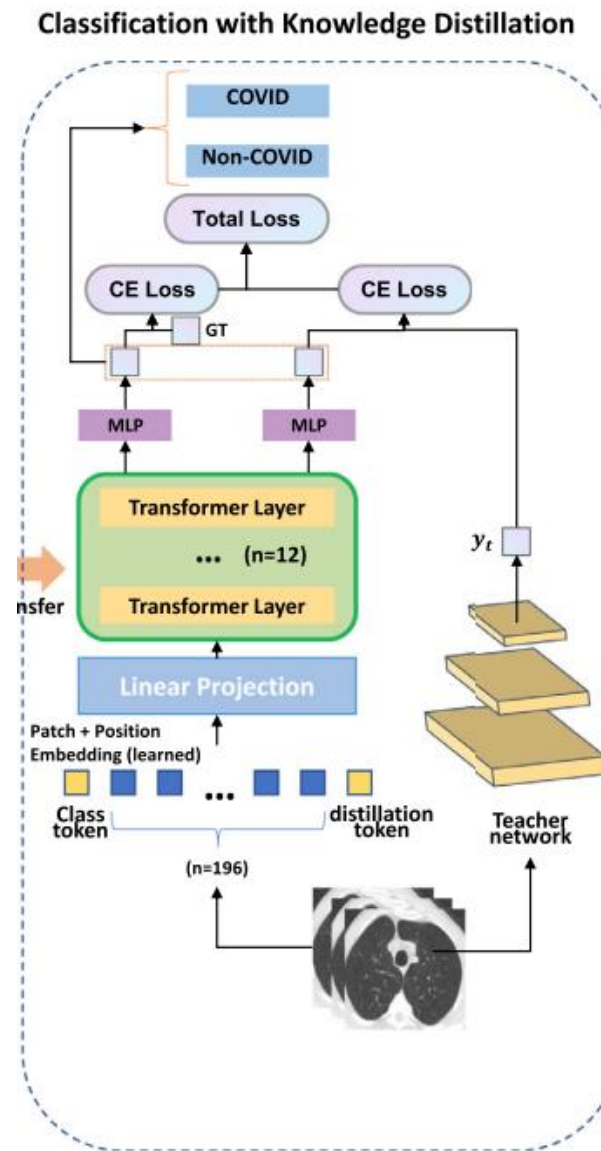
Dlatego sieci konwolucyjne są użyte jako swoisty nauczyciel dla ViT. Pozwala to ViT lepiej koncentrować się na lokalnych cechach COVID-19 podczas ekstrakcji cech globalnych.

# KLASYFIKACJA Z DESTYLACJĄ WIEDZY

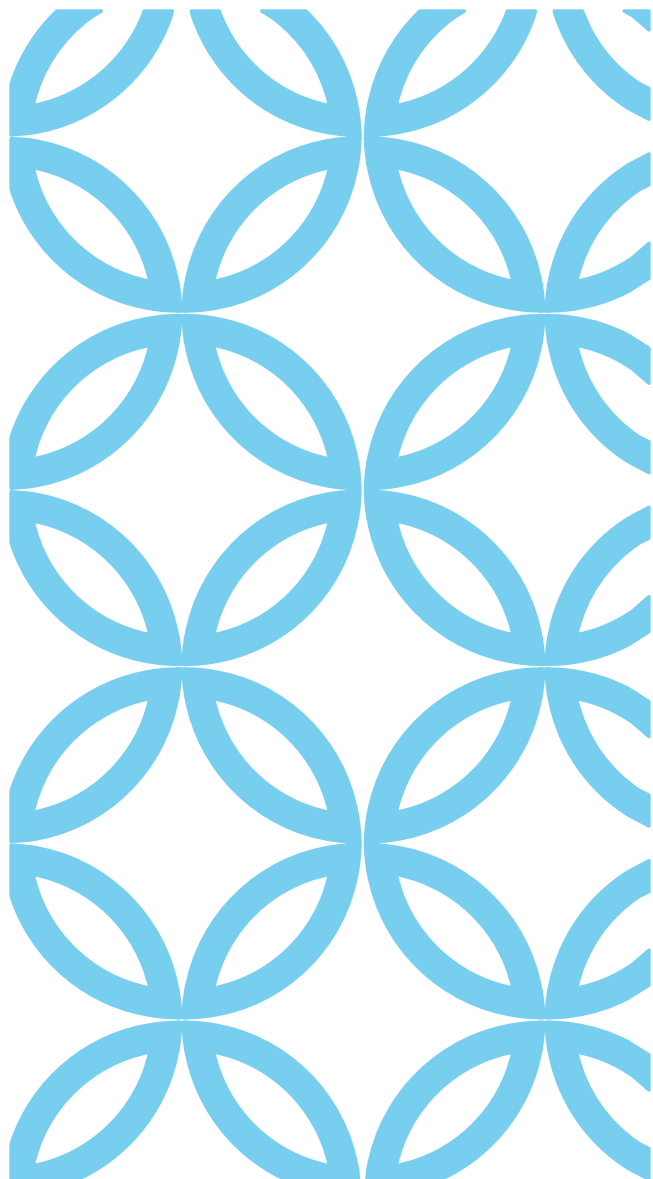


# KLASYFIKACJA Z DESTYLACJĄ WIEDZY

Dodany jest nowy token destylacji, który współdziela z tokenem klasy i tokenem fragmentu przez warstwę samo-uwagi i jest prezentowany na końcowej warstwie połączonej. Token destylacji działa podobnie do tokena klasy. Jedyna różnica polega na tym, że celem tokena klasy jest zgodność z rzeczywistą etykietą, podczas gdy token destylacji ma być zgodny z etykietą przewidzianą przez model nauczyciela (CNN).







---

# EKSPERYMENTY

# WYNIKI EKSPERYMENTÓW

Do ewaluacji modelu, są użyte takie metryki:

- Recall
- Precision
- Accuracy
- AUC
- F1-Score

**Table 2**Classification results of different methods on two COVID-19 CT datasets (**bold**: best).

	COVID-19 CT Dataset					SARS-Cov-2 CT Dataset				
	Recall	Precision	Accuracy	F1	AUC	Recall	Precision	Accuracy	F1	AUC
ResNet18	44.90±1.34	81.48±1.79	68.47±0.48	57.89±2.59	83.53±1.04	69.01±1.69	<b>100.00± 0.00</b>	84.35±0.48	81.66±1.37	95.25±1.36
ResNet50	42.86±1.23	87.50±0.17	69.46±1.29	57.53±2.68	81.42±2.05	79.23±1.10	99.59± 0.07	89.35±1.15	88.26±1.27	96.47±1.04
DenseNet121	83.67± 2.61	82.00±1.37	83.25±1.82	82.82±1.74	88.63±1.86	79.55±1.09	99.60±0.14	89.52±1.19	88.45±0.52	95.70±0.87
DenseNet169	57.14±2.71	91.80± 1.06	76.84±1.59	70.44±0.96	87.02±2.42	80.83±1.67	99.61± 0.17	90.16±1.01	89.24±1.49	97.83± 1.27
VGG16	52.05±3.12	82.26±1.89	71.43±2.29	63.75±1.94	83.31±0.58	80.51±2.46	98.82±0.71	89.68±1.08	88.73±1.37	97.82±0.74
ConvNeXt [25]	77.75±1.15	66.09±1.04	69.95±1.28	71.36±0.82	75.07±0.78	72.84±1.02	95.80±0.82	84.68±1.25	82.76±1.75	92.78±2.34
COVID-Net [26]	57.73±2.94	64.03±3.91	63.12±2.09	61.09±1.28	71.09±2.18	70.97±2.37	80.04±2.87	77.12±0.98	76.03±1.13	84.08±0.92
COVID-Net-R [27]	74.69±3.91	79.48±0.96	77.07±1.92	77.04±2.17	84.13±0.82	83.78±0.62	94.58±2.07	89.09±1.08	88.97±0.91	94.12±0.87
Series Adapter [28]	74.91±1.89	63.04±4.87	70.01±3.82	67.08±3.09	73.92±2.36	81.91±2.61	90.08±0.79	85.73±0.71	86.19±1.65	92.93±1.42
Parallel Adapter [29]	71.81±2.47	79.84±1.75	74.93±1.83	73.46±1.68	80.29±1.76	80.02±2.47	83.51±1.87	82.13±1.91	82.39±1.78	89.99±0.97
MS-Net [30]	74.07 ±1.29	79.29±1.48	76.23±1.81	76.54±1.73	82.19±1.47	84.91±2.83	93.78±2.76	87.98±1.31	88.73±1.20	94.37±0.79
SN-Contrast [27]	79.71±1.42	78.02±1.34	78.69±1.54	78.83±1.43	85.32±0.32	85.89±1.05	95.75±0.43	90.83±0.93	90.87±1.29	96.24±0.35
Ensemble DCNN [31]	<b>84.97± 2.30</b>	85.54±2.52	84.97±2.30	85.03± 2.35	89.48± 2.42	95.57± 0.51	95.59±0.52	95.57± 0.51	95.57± 0.51	97.94± 0.83
SwinT V1 [32]	70.41±1.71	80.23±0.79	77.34±1.92	75.00±1.35	85.35±2.16	80.83±2.24	97.68±0.48	89.35±0.84	88.64±1.85	94.77±1.34
SwinT V2 [33]	66.33±1.29	87.84± 2.83	79.31±0.34	75.58±1.94	86.77±2.59	81.79±0.58	95.88±1.46	89.03±3.58	88.28±1.52	95.39±1.42
DAT [34]	70.41±2.28	74.19±1.59	73.89±2.21	72.25±2.36	80.44±1.36	82.43±1.18	96.27±1.20	89.52±1.58	88.81±0.78	96.55±1.42
Conformer [35]	79.59±2.53	86.67±0.56	84.24± 1.92	82.98± 2.53	89.69± 1.43	90.42± 2.59	94.65±1.46	92.58± 1.35	92.48± 1.45	95.32±1.41
EViT [36]	52.04±3.28	76.12±2.42	68.97±1.69	61.82±2.65	74.11±1.83	81.15±2.71	96.21±1.64	88.87±1.06	88.04±2.20	94.89±1.06
PCPVT-B [37]	67.35±1.41	84.62±2.52	78.33±1.53	75.00±3.12	86.60±1.35	83.71±1.42	96.32±0.81	90.16±1.38	89.57±2.10	95.22±0.89
PVT V1 [38]	64.29±2.16	75.90±1.34	72.91±1.53	69.61±0.96	76.33±1.57	82.75±1.98	94.87±1.04	89.03±1.77	88.40±2.10	94.13±1.31
PVT V2 [39]	79.59±1.54	82.98±2.04	82.27±0.79	81.25±2.05	85.88±1.42	83.39±1.64	97.75±1.32	90.65±2.45	90.00±1.38	95.43±1.11
Ours (MCT-KD)	81.63± 1.64	<b>94.12± 1.26</b>	<b>87.43± 1.32</b>	<b>88.67± 1.12</b>	<b>92.23± 0.54</b>	<b>96.17± 0.48</b>	97.73±1.34	<b>96.94± 0.73</b>	<b>96.94± 1.03</b>	<b>99.60± 0.38</b>

Porównujemy naszą metodę MCT-KD z tradycyjnymi metodami uczenia głębokiego, takimi jak ResNet18, ResNet50, DenseNet121, DenseNet169 i VGG16. W tabeli widać, że DenseNet121 dobrze radzi sobie z zestawem danych COVID-19 CT, ale MCT-KD ma lepszą wydajność. Precision metody MCT-KD jest o 12,12% wyższa niż DenseNet121, co wskazuje na niskie ryzyko false positive wyników. Chociaż tradycyjne metody mają lepszą precyzję na zestawie danych SARS-Cov-2 CT, MCT-KD znacząco poprawia recall, wskazując, że podstawowe metody nie są skuteczne w identyfikacji pozytywnych pacjentów COVID-19. W diagnozowaniu COVID-19 MCT-KD ma przewagę nad tradycyjnymi metodami uczenia głębokiego.





Porównujemy naszą metodę MCT-KD z obecnymi najlepszymi metodami klasyfikacji COVID-19: metodami end-to-end, wieloetapowymi i opartymi na Transformerach. Wszystkie metody były testowane na tym samym zestawie danych.

Metody end-to-end: Wśród nich są ConvNeXt, COVID-Net i COVID-Net-R. MCT-KD osiąga znacząco lepsze wyniki, w tym 11,63% większą dokładność i 14,64% większą precyzję niż COVID-Net-R na danych COVID-19 CT.

Metody wieloetapowe: Takie jak Series Adapter, Parallel Adapter i inne. MCT-KD ma nieco niższy recall niż Ensemble DCNN, ale 8,58% wyższe accuracy na zestawie COVID-19 CT.

Metody oparte na Transformerach: Takie jak SwinT V1, DAT i inne. MCT-KD wykazuje znaczącą przewagę, zwłaszcza pod względem accuracy. Warto zaznaczyć, że Conformer integruje cechy globalne i lokalne, co skutecznie klasyfikuje COVID-19.

Podsumowując, w porównaniu z wieloma nowoczesnymi metodami, MCT-KD wypada bardzo korzystnie w większości metryk. Przyczyną tego jest kombinacja cech globalnych (bazując na MCT) i lokalnych z CNN poprzez destylację wiedzy.



**Table 3**

Classification results of different methods on CAP and COVID-19 Dataset (**bold**: best).

	Recall	Precision	Accuracy	F1	AUC
ResNet50	78.61±1.23	87.57±0.17	81.65±1.29	82.85±2.68	90.22±2.05
DenseNet121	82.63±2.61	76.09±1.37	75.57±1.82	79.22±1.74	82.20±1.86
DenseNet169	<b>94.51± 2.71</b>	74.28±1.06	78.45±1.59	83.19±0.96	85.60±2.42
VGG16	60.32±3.12	97.06± 1.89	76.60±2.29	74.41±1.94	87.13±0.58
ConvNeXt [25]	83.55±1.15	67.11±1.04	67.63±1.28	74.43±0.82	63.92±0.78
Conformer [35]	79.34±1.36	91.18±1.64	84.02±2.14	84.85±0.72	92.94±1.57
EViT [36]	89.40± 1.38	83.16±0.96	83.81±1.70	86.17±1.02	88.77±1.48
PCPVT-B [37]	91.59± 1.48	83.50±1.57	85.05± 2.54	87.36± 2.82	92.42±1.48
PVT V1 [38]	87.02±1.65	85.77±1.52	84.54± 1.96	86.39±2.51	93.67± 1.35
PVT V2 [39]	78.43±1.32	<b>97.72± 1.63</b>	82.27±1.28	86.80± 2.83	94.48± 1.93
Ours (MCT-KD)	82.45±1.28	96.78± 1.05	<b>88.56± 1.27</b>	<b>89.04± 2.38</b>	<b>95.73± 1.06</b>

Wybieramy kilka metod opartych na CNN i Transformerze do porównania. Jak pokazuje tabela, metody oparte na transformerach generalnie przewyższają te oparte na CNN. MCT-KD wciąż ma przewagę w trzech metrykach. Jednak te przewagi osłabły w porównaniu z poprzednim zadaniem klasyfikacyjnym i wystąpił niewielki spadek accuracy, który prawdopodobnie wynika z podobieństwa między CAP a COVID-19.



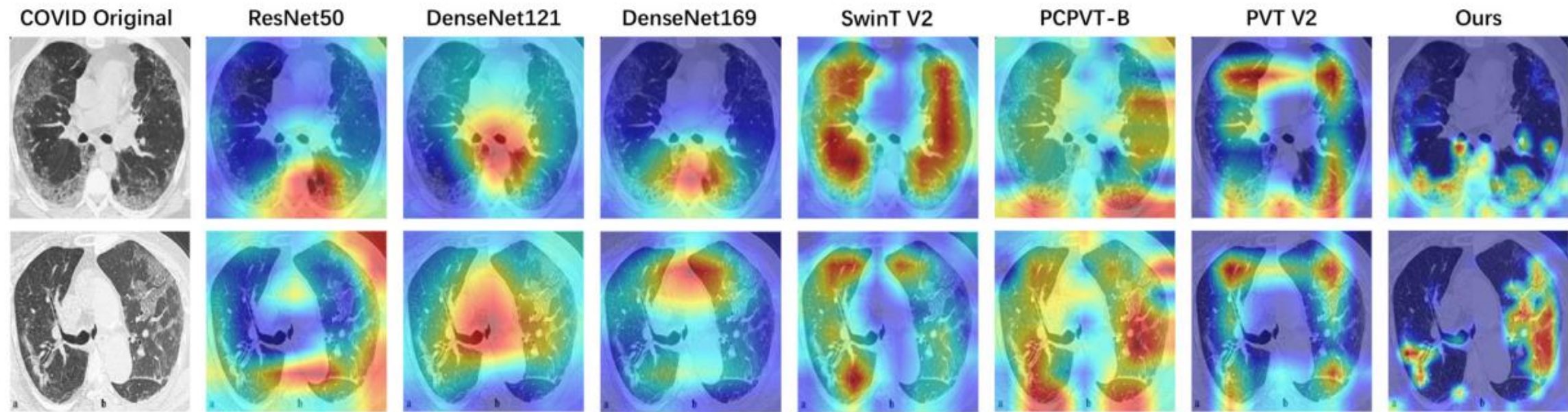
**Table 4**Effects of pre-training on downstream classification task (**bold**: best).

	Recall	Precision	F1	Accuracy	AUC
Non-pre-training	74.94 $\pm$ 1.43	88.25 $\pm$ 2.04	81.05 $\pm$ 1.28	82.50 $\pm$ 1.10	89.73 $\pm$ 2.24
Ours (MCT-KD)	<b>86.86<math>\pm</math> 1.28</b>	<b>95.57<math>\pm</math> 1.05</b>	<b>91.19<math>\pm</math> 0.92</b>	<b>91.62<math>\pm</math> 1.34</b>	<b>96.68<math>\pm</math> 1.84</b>

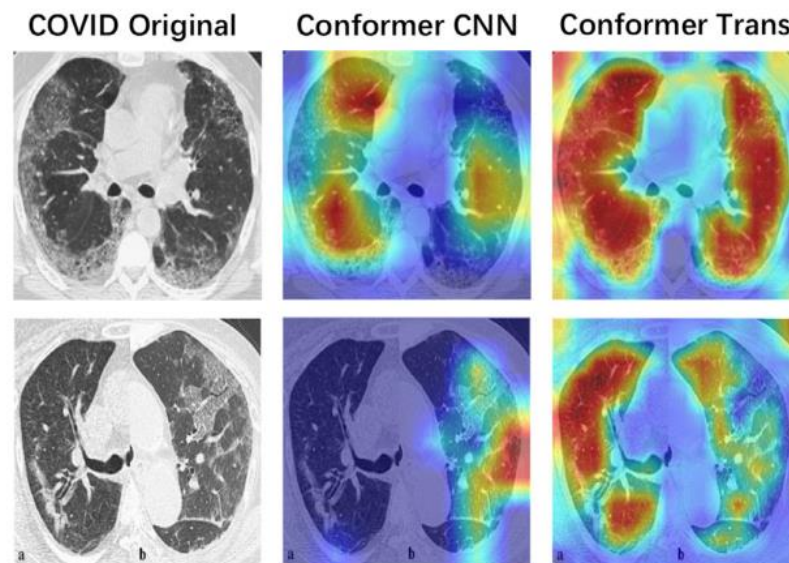


Aby zademonstrować skuteczność reprezentacji cech bazujących na MCT, porównujemy wyniki przed i po transferze. Dla szybkości szkolenia łączymy zestawy danych COVID-19 i SARS-Cov-2 w jeden zestaw do kolejnych eksperymentów. Wyniki w tabeli pokazują, że wydajność MCT-KD znacząco wzrosła dzięki użyciu reprezentacji cech opartej na MCT.





**Fig. 5.** Visualization of features using Grad-CAM [40]. It is a class-discriminative localization technique that generates visual explanations for deep learning model decisions. The heat map obtained by using Grad-CAM to visualize the features of different methods shows that MCT-KD pays more attention to the lesion area when making decisions. This can provide a basis for clinical diagnosis.



**Fig. 6.** Visualization for CNN and Transformer branch of Conformer [35].



Aby zilustrować interpretowalność proponowanej metody MCT-KD, używamy mapowania aktywacji klasy ważonej gradientem (Grad-CAM) do podkreślenia ważnych obszarów na obrazach RTG klatki piersiowej związanych z podejmowaniem decyzji. Wynik jest pokazany na rysunku 5. Pierwsza kolumna przedstawia oryginalne obrazy RTG klatki piersiowej COVID-19, a ostatnia naszą proponowaną metodę w porównaniu z innymi. W porównaniu z innymi mapami cieplnymi, mapy uzyskane przez MCT-KD pokrywają więcej cech choroby, co stanowi podstawę do poprawnej diagnozy COVID-19.



# KONKLUZJA

Proponujemy metodę MCT-KD do analizy obrazów RTG klatki piersiowej COVID-19, poprawiając działanie ViT na małych zestawach danych. Używając samo-nadzoru i destylacji wiedzy, metoda umożliwia dokładniejszą diagnozę. Zastosowaliśmy Grad-CAM do wizualizacji uwagi modelu na obszarze zmian. Mimo pewnych zalet MCT-KD, jej dokładność w pewnych zadaniach nie jest satysfakcjonująca, co planujemy badać w przyszłości.





DZIĘKUJEMY