

Sztuczna Inteligencja w Grafice Komputerowej - Projekt nr 3

Błażej Ejzak, Paweł Dombrzalski

313220, 318647

9 maja 2025

Streszczenie

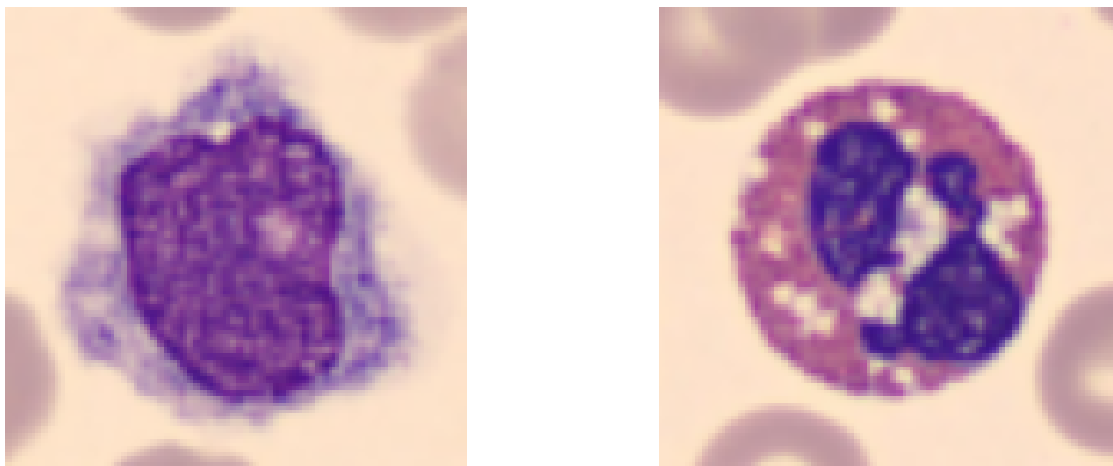
Niniejszy dokument prezentuje wyniki rozwiązania zadania związanego z generatywnymi sieciami neuronowymi.

Spis treści

1	Charakterystyka zbioru danych	2
2	VAE	2
3	GAN	4
4	Dyfuzja	7

1 Charakterystyka zbioru danych

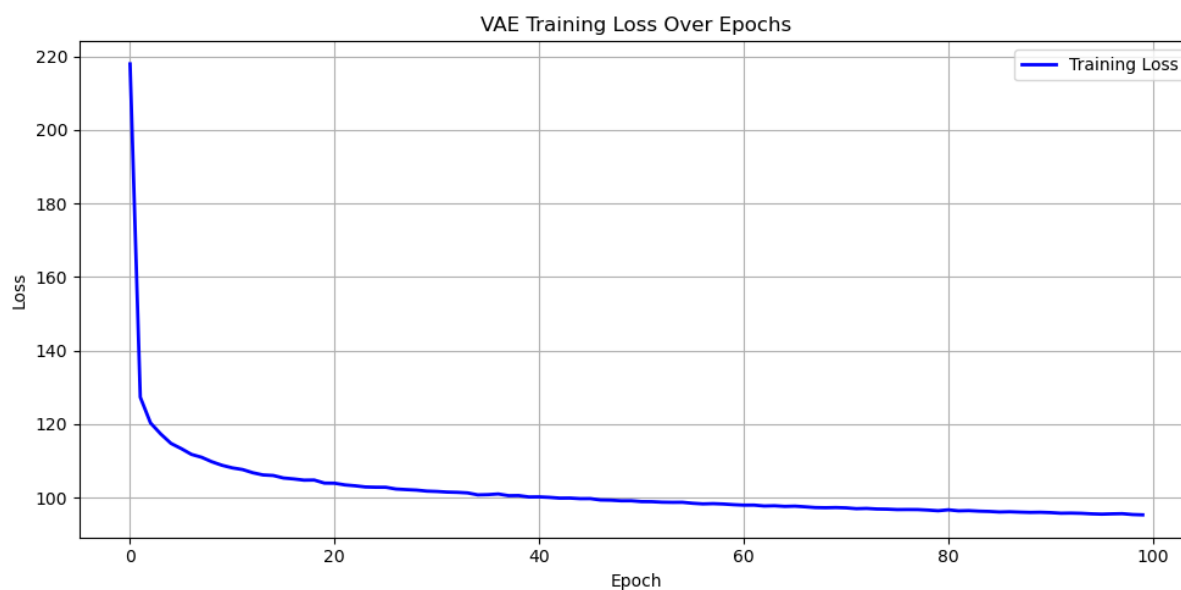
Jako podstawę do uczenia oraz porównywania wyników wybraliśmy zbiór BloodMNIST przedstawiający strukturę mikroskopową osocza człowieka.



Rysunek 1: Wybrane zdjęcia ze zbioru BloodMNIST

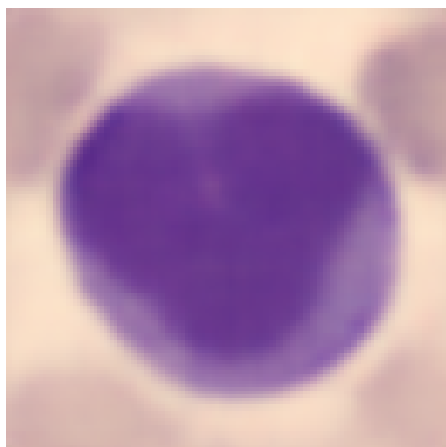
2 VAE

Wariacyjny Autoenkoder to model który składa się z enkodera, który koduje do przestrzeni opisanymi rozkładami gaussa (średnia oraz odchylenie standardowe) oraz dekodera, który odwraca ten proces. Dekoder oraz enkoder są modelami konwolucyjnymi. Generacja odbywa się poprzez wylosowanie punktu/wektora z przestrzeni ukrytej, a następnie przekształcenie tego wektora za pomocą dekodera. Proces uczenia gdzie funkcją straty jest $MSE + KLD$ przedstawiony jest na poniższym wykresie:

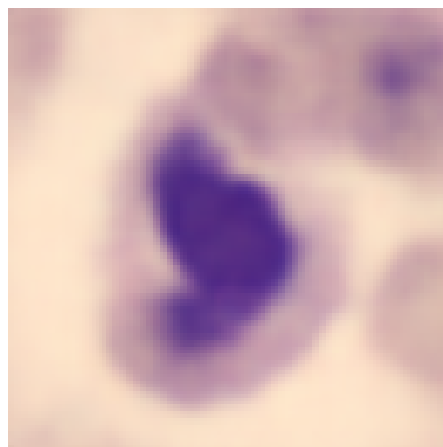


Rysunek 2: Wyniki uczenia VAE

Pred: 5



Pred: 6

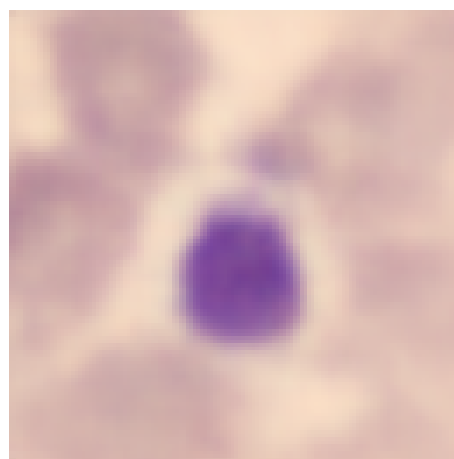


Rysunek 3: VAE - Wygenerowane zdjęcia

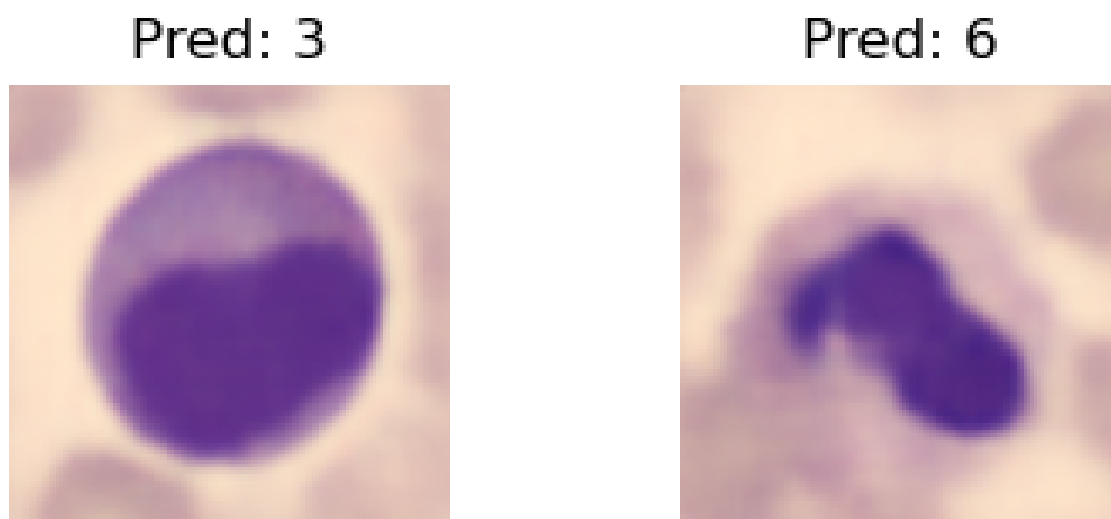
Pred: 6



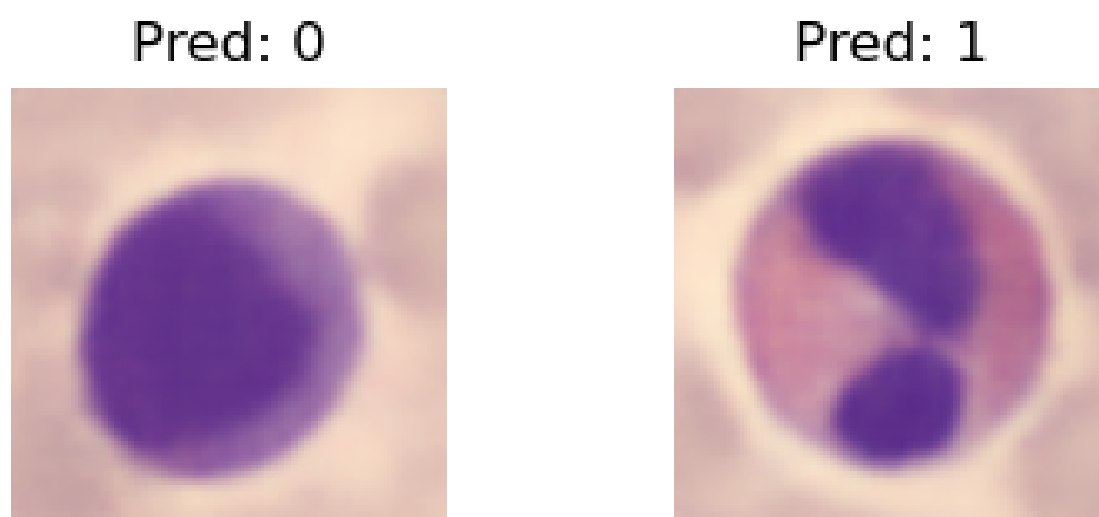
Pred: 7



Rysunek 4: VAE - Wygenerowane zdjęcia



Rysunek 5: VAE - Wygenerowane zdjęcia



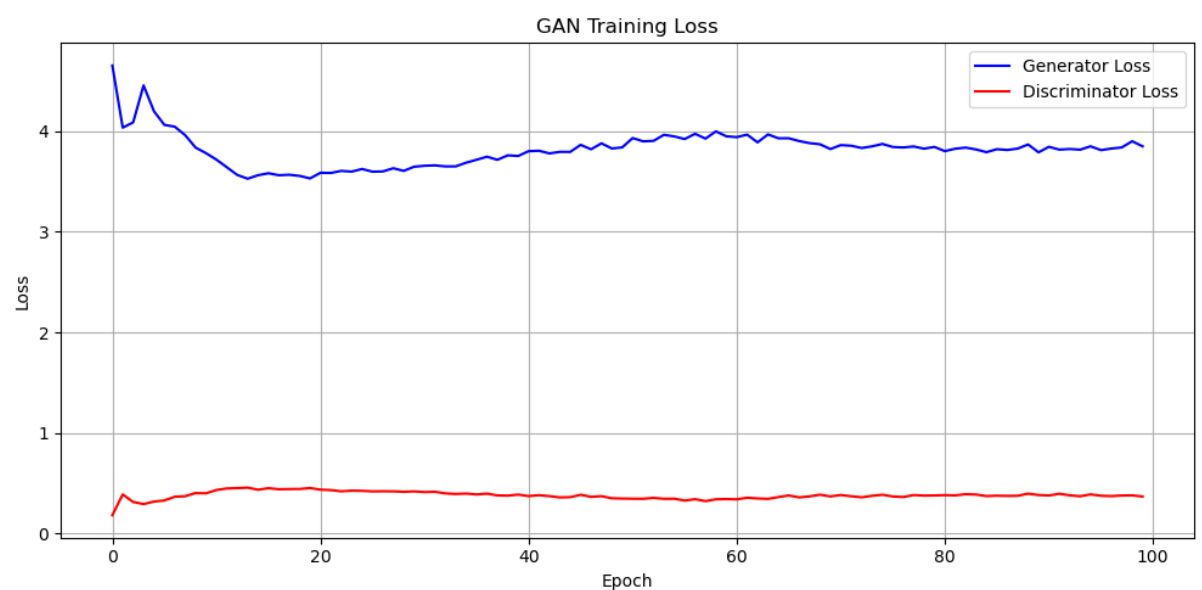
Rysunek 6: VAE - Wygenerowane zdjęcia

3 GAN

GAN to model, który składa się z dwóch zasadniczych części:

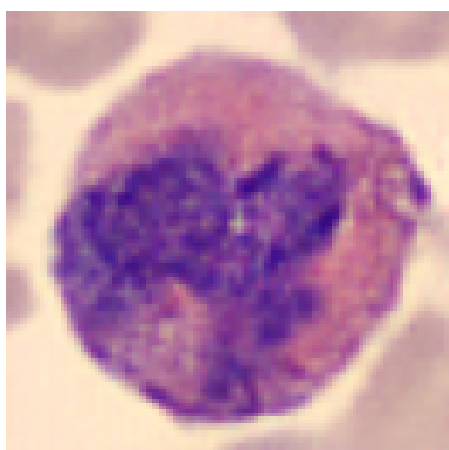
- Generator będący konwolucyjną siecią neuronową, który stara się oszukać dyskryminator (w procesie uczenia wykorzystuje adversarialną funkcję straty).
- Dyskryminator będący klasyfikatorem binarnym odróżniający obrazy oryginalne od wygenerowanych (w procesie uczenia wykorzystuje entropię krzyżową).

Na poniższym wykresie przedstawione zostały efekty uczenia się modelu GAN:

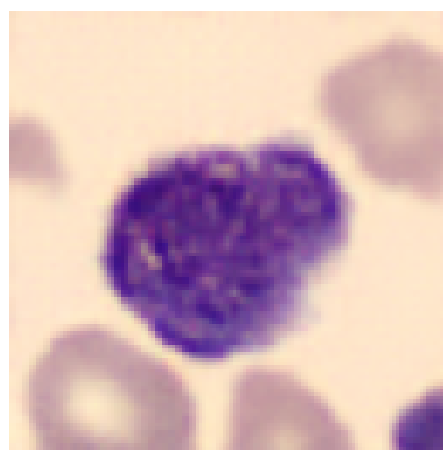


Rysunek 7: Wyniki uczenia GAN

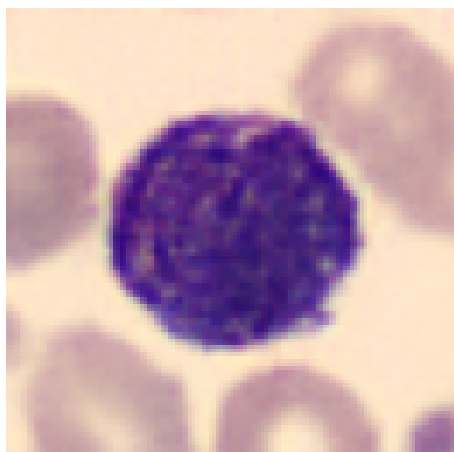
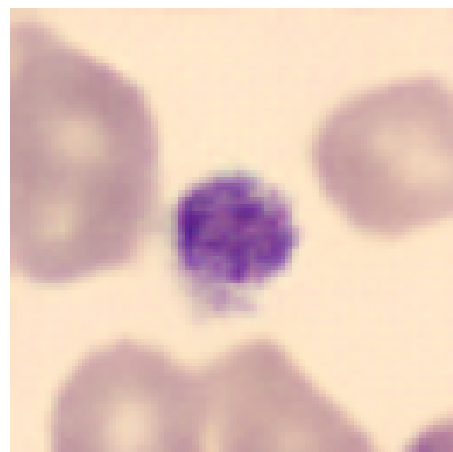
Pred: 4



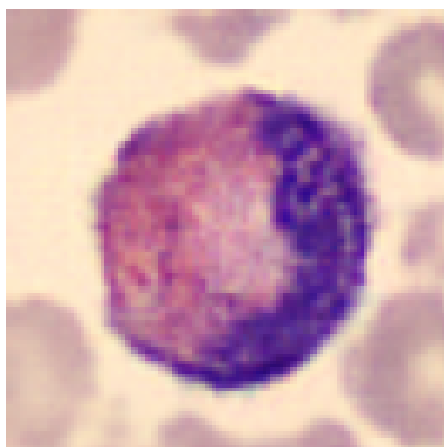
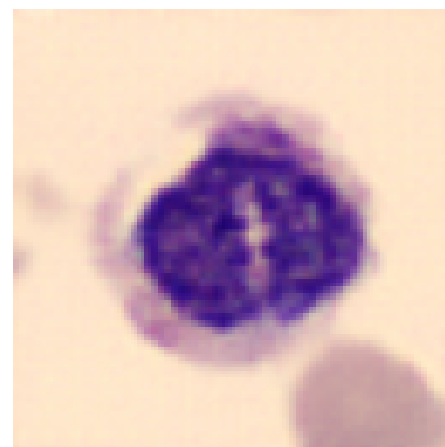
Pred: 3



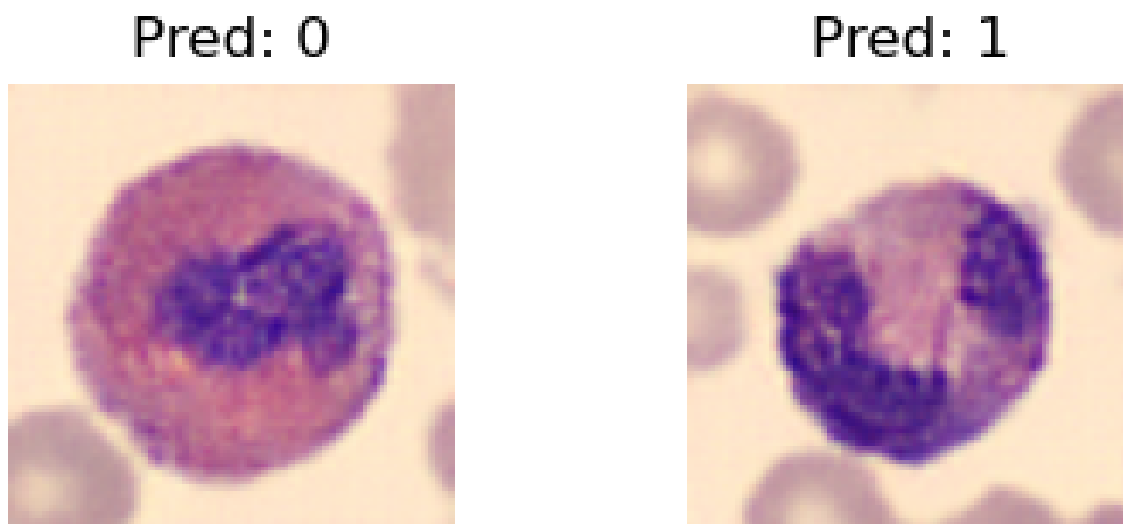
Rysunek 8: GAN - Wygenerowane zdjęcia

Pred: 6**Pred: 5**

Rysunek 9: GAN - Wygenerowane zdjęcia

Pred: 3**Pred: 4**

Rysunek 10: GAN - Wygenerowane zdjęcia



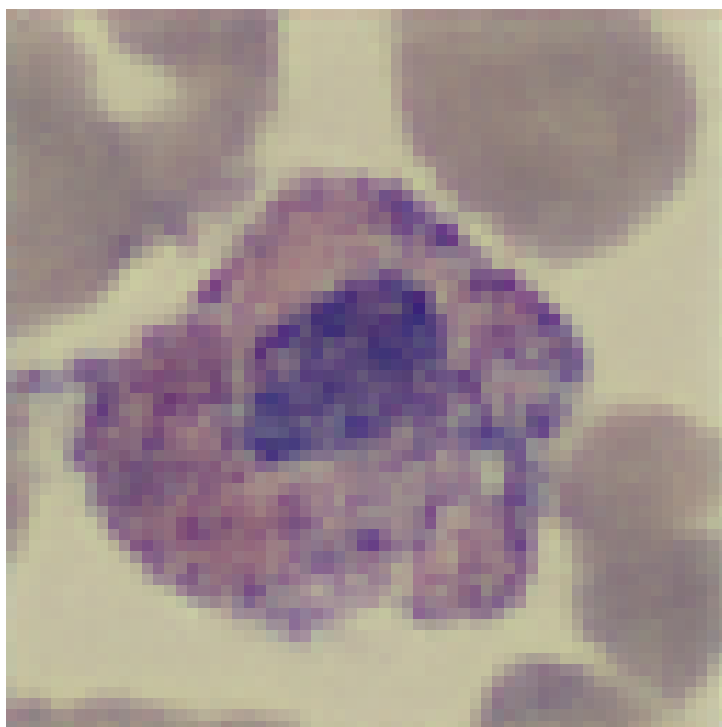
Rysunek 11: GAN - Wygenerowane zdjęcia

4 Dyfuzja

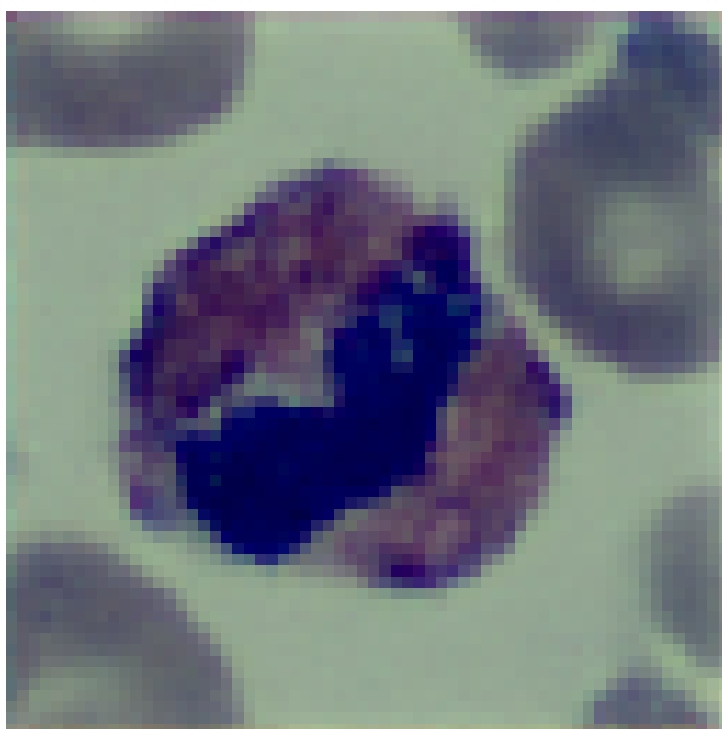
Ostatnim zaimplementowanym modelem był model dyfuzyjny. Generacja obrazów polega na stopniowym odszumianiu z losowego szumu Gaussa. Model ma zdefiniowane 2 procesy - *forward process* i *reverse process*. *Forward process* stosowany jest podczas treningu i polega na stopniowym dodawaniu szumu do obrazu przez określoną liczbę kroków t - np. dla 1000 kroków, obraz przy $t = 0$ nie ma żadnego nałożonego szumu Gaussa, natomiast dla $t = 1000$ obraz jest w całości losowym szumem. Przy treningu model uczy się stopniowo usuwać szum z obrazu - do tego wykorzystano sieć o architekturze U-Net. Liczony jest błąd średniokwadratowy między szumem przewidzianym przez U-Net, a rzeczywiście dodanym do obrazu. Podczas generacji z losowego szumu, U-Net przewiduje przez T kroków (w tym przypadku 1000) szum na obrazie, który jest usuwany poprzez *reverse process*. Aby umożliwić generację obrazów o specyficznych klasach, dodano *embedding* klasy.

Eksperyment	Dokładność
Zbiór mieszany	0.475
Trening oryginalny, test wygenerowany	0.129
Trening wygenerowany, test oryginalny	0.137

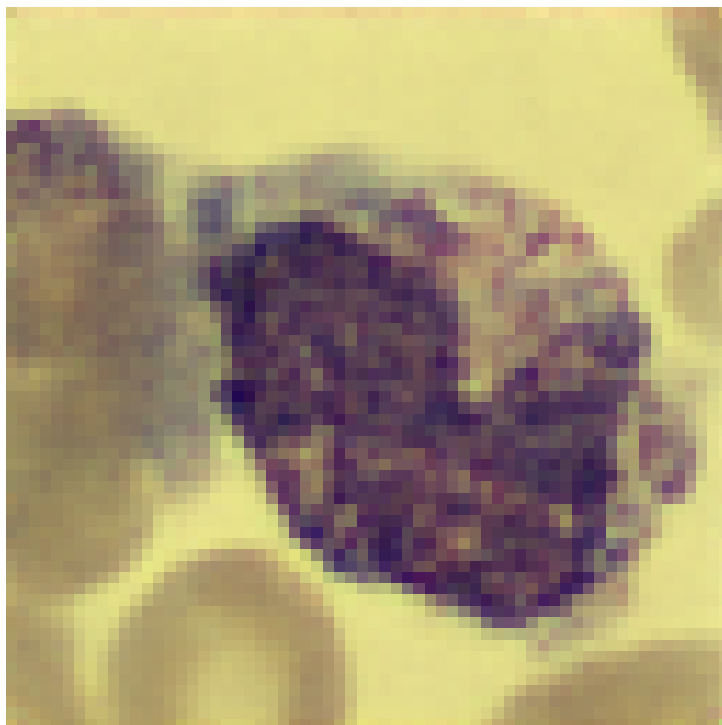
Ostatecznie obrazy generowane przez model dyfuzyjny nie dały dobrych wyników - co wynikało z faktu, że często model nakładał filtr koloru na wyjściowe zdjęcie. Mogło to wynikać ze zbyt płytkiej architektury U-Net - o zbyt małej liczbie *skip connections*. Kolejną przyną były ograniczenia w ilości wygenerowanych zdjęć - w porównaniu z VAE czy GANem, modele dyfuzyjne mają większą złożoność obliczeniową (przy generacji obrazu potrzebne jest przejście przez 1000 kroków procesu odszumiania), co znacząco wydłuża czas generacji. Przyczyniło się to do mniejszej ilości zdjęć podczas treningu klasyfikatora - co pogorszyło wyniki.



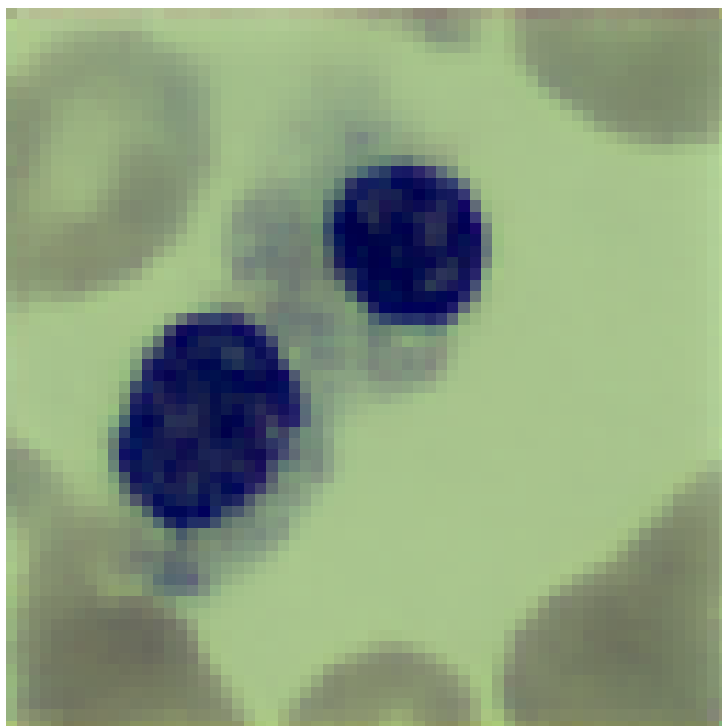
Rysunek 12: Dyfuzja - Wygenerowane zdjęcie - klasa 1



Rysunek 13: Dyfuzja - Wygenerowane zdjęcie - klasa 1



Rysunek 14: Dyfuzja - Wygenerowane zdjęcie - klasa 4



Rysunek 15: Dyfuzja - Wygenerowane zdjęcie - klasa 4

Klasa	Dokładność	Liczba obrazów
0	0.00	677
1	0.91	2194
2	0.97	1343
3	0.14	1793
4	0.21	616
5	0.31	933
6	0.88	3062
7	0.77	1382
Ogółem	0.64	12000

Tabela 1: VAE - Dokładność oraz liczba otagowanych zdjęć (oryginalne - wygenerowane)

Klasa	Dokładność	Liczba obrazów
0	1.00	852
1	0.92	2181
2	0.83	1085
3	0.46	2026
4	0.99	849
5	0.97	993
6	0.77	2330
7	1.00	1643
Ogółem	0.83	11959

Tabela 2: VAE - Dokładność oraz liczba otagowanych zdjęć (wygenerowane - oryginalne)

Klasa	Dokładność	Liczba obrazów
0	0.86	1529
1	0.95	4375
2	0.95	2428
3	0.78	3819
4	0.93	1465
5	0.97	1926
6	0.81	5392
7	0.85	3025
Ogółem	0.87	23959

Tabela 3: VAE - Dokładność oraz liczba otagowanych zdjęć (zmieszane zbiory wygenerowane i oryginalne)

Klasa	Dokładność	Liczba obrazów
0	0.55	1252
1	0.58	2316
2	0.48	297
3	0.51	3334
4	0.52	326
5	0.57	857
6	0.50	1135
7	0.60	2483
Ogółem	0.54	12000

Tabela 4: GAN - Dokładność oraz liczba otagowanych zdjęć (oryginalne - wygenerowane)

Klasa	Dokładność	Liczba obrazów
0	0.60	852
1	0.70	2181
2	0.65	1085
3	0.68	2026
4	0.66	849
5	0.64	993
6	0.59	2330
7	0.66	1643
Ogółem	0.65	11959

Tabela 5: GAN - Dokładność oraz liczba otagowanych zdjęć (wygenerowane - oryginalne)

Klasa	Dokładność	Liczba obrazów
0	0.98	2104
1	0.68	4497
2	1.00	1382
3	0.46	5360
4	0.99	1175
5	0.98	1850
6	0.89	3465
7	0.50	4126
Ogółem	0.81	23959

Tabela 6: GAN - Dokładność oraz liczba otagowanych zdjęć (zmieszane zbiory wygenerowane i oryginalne)