

AI w parametryzacji danych oceanograficznych

Aleksandra Justyna Kaleta

Uniwersytet Gdański, Wydział Oceanografii i Geografii

E-mail: a.kaleta.084@studms.ug.edu.pl

tutor: dr hab. Marcin Paszkuta

Uniwersytet Gdański, Wydział Oceanografii i Geografii

Katedra Geofizyki

Słowa kluczowe: nauczanie maszynowe, analiza danych niestabilnych, oceanografia, lodowce

Streszczenie

Nauczanie maszynowe jest dynamicznie rozwijającą się częścią sztucznej inteligencji. W ostatnich latach coraz częściej zaczynamy korzystać z jego nieograniczonych zastosowań. Celem publikacji jest przedstawienie możliwości wykorzystania sztucznej inteligencji w analizach niestabilnych danych oceanograficznych, gdzie klasyczne metody nie dają satysfakcjonujących wyników. Wykonano nienadzorowaną parametryzację oraz przedstawiono przykłady eksperymentalne serii danych: satelitarnych NASA i arktycznych ekspedycji oceanograficznych. Proponowana metoda, opierająca się na binaryzacji oraz parametryzacji, działa dobrze w przypadku klasyfikacji lodu oraz miejsc potencjalnego cielenia lodowców. Dużą rolę odgrywa właściwe ustalenie progu binaryzacji i jest to potencjalna rola nauczania maszynowego.

Obok przykładów w formie graficznej przedstawiono podstawowe statystyki określające m.in. zmiany wartości entropii.

Wstęp

Nauczanie maszynowe (ang. *Machine Learning*) jest częścią sztucznej inteligencji (AI-ang. *Artificial Intelligence*), która polega na rozwijaniu umiejętności komputerów, poprzez analizy nowych i wcześniejszych doświadczeń. Dzięki temu nie istnieje potrzeba bezpośredniego programowania [1]. Algoritmy używane w nauczaniu maszynowym bardzo dobrze radzą sobie z dużą ilością danych (ang. *big data*) – obowiązuje zasada, że im więcej danych tym sprawniej system jest w stanie się nauczyć. Bez względu na to jak dużą ilość danych dysponujemy muszą one posiadać pewne zależności między sobą, te zależności pozwalają sprawniej zrozumieć informacje (Alpaydin, 2021). Potencjał nauczania maszynowego jest bardzo szeroki i daje możliwości zastosowania w życiu codziennym, w nauce (Woodman i Mangoni, 2023; Ekasari i in., 2024; Ponomarenko i in.,

2024) jak i edukacji (Paszruta, 2023; Paszruta i Szymczak, 2024).

Nauczanie maszynowe znajduje swoje zastosowanie również w naukach o Ziemi m.in. przy parametryzacji danych oceanograficznych tj.: przewidywaniu zmian klimatu oceanicznego, identyfikacji gatunków, modelowaniu falowania i wiatru, wykrywaniu plam olejowych oraz wielu innych (Ahmad, 2019). Dryfujący lód jak i lodowce są ściśle związane z warunkami klimatycznymi i dynamiką wód, a co za tym idzie ich monitoring przy pomocy technik nauczania maszynowego umożliwia przewidywanie zmian środowiska (Ahmad, 2019). Regiony polarne są jednymi z najbardziej narażonych na zmiany cieplarniane. Jedną z konsekwencji ocieplenia klimatu jest topnienie lodowców oraz zmiany pokrywy lodowej (Stocker i in., 2020). Wiele fragmentów lodu odłamywanych od lodowców lub pozostałych po roztopieniu pokrywy lodowej dryfuje na powierzchni wody stanowiąc zagrożenie dla poruszających się po niej statków. Znacznym zagrożeniem dla bezpieczeństwa jest również nagłe odrywanie się dużych części lodowca. Często trudne do przewidzenia jest to która część lodowca jest bardziej podatna na destabilizację, co również stanowi niebezpieczeństwo związane z eksplorowaniem tych rejonów (Duvillard i in., 2015; Faillettaz i in., 2015; Haeberli i in., 2017; Marcer i in., 2019; Zakhvatkina i in., 2019).

Celem publikacji jest zastosowanie nauczania maszynowego do parametryzacji danych oceanograficznych związanych bezpośrednio z bezpieczeństwem środowiska, mającym na celu jego ochronę przed zniszczeniem do stopnia zagrażającego ludziom (Molvaer, 1991). Osiągnięte jest to głównie przez monitorowanie zmian klimatu. Przedstawiono eksperyment mający za zadanie identyfikację: lodu na zdjęciach satelitarnych NASA (*National Aeronautics and Space Ad-*

ministration) oraz obszarów cielenia lodowców na podstawie obrazów z wypraw polarnych (z wykorzystaniem wysokiego oświetlenia powierzchni). Wybór powyższych danych został dokonany ze względu na trudności przy parametryzacji tradycyjnymi metodami.

Materiały i metody

Eksperyment został podzielony na dwie części, w każdej z nich analizowane były dane oceanograficzne przy użyciu tego samego kodu. Do zadań nauczania maszynowego należało wyznaczenie właściwego progu binaryzacji, dodatkowo przed jak i po analizie, obliczana była entropia. Entropię obliczano z danych, które były przetwarzane w formacie graficznym PNG. Prace prowadzono w środowisku OCTAVE z bibliotekami AI. Do pierwszej części eksperymentu wykorzystano serię danych NASA Earth Observatory [2] na których widoczny jest lód, często również odłamki lodu oraz woda, w niektórych przypadkach obszary nieba i lądu pokrytego śniegiem. W drugiej części wykorzystano zdjęcia z wypraw polarnych, na których widoczne są lodowce z perspektywy horyzontalnej. Oprócz samego lodowca na zdjęciach często również występują obszary nieba z chmurami, lądu i wody.

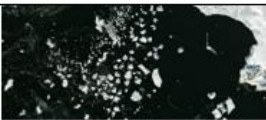
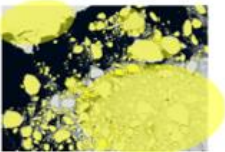
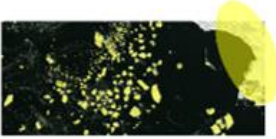
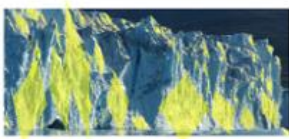
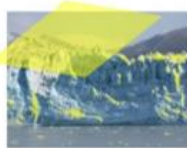
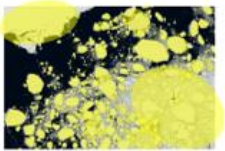
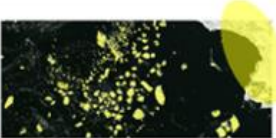
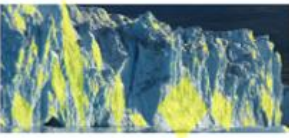
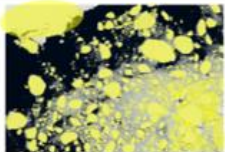
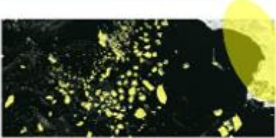
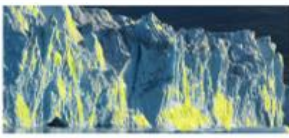
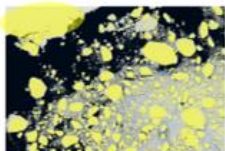
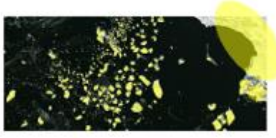
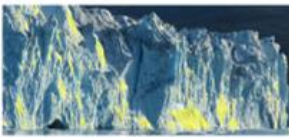
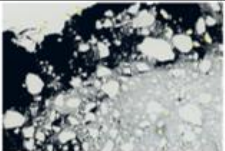
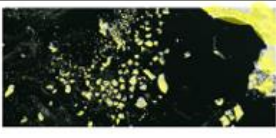
Na potrzeby eksperymentu stworzony został skrypt. W początkowym etapie jego działania dane, w formie zdjęć, zostają wczytane jako macierz, która następnie jest binaryzowana z progiem określanym przez algorytm AI. Próg ten znajduje się w granicach 0-1 i określany jest dla najlepiej reprezentowalnych wyników doświadczeń, czyli takich, w których najtrafniej rozpoznane zostały odłamywane lub naświetlone części lodowca, w zależności od części eksperymentu. Wyniki te zostały przedstawione w dalszej części pracy. Uzyskane parametry reprezentujące ilość klasyfikowanych elementów, długości osi głównej i krótkiej każdego z elementów

skalsyfikowanych oraz ich orientacje wyznaczono na podstawie największej sumy długości osi głównej i pomocniczej. Na potrzeby eksperymentu kształt analizowanego obiektu definiowany był jako elipsa lub czworokąt. W kolejnym kroku zdefiniowane kształty zostały wyświetlone na tle klasyfikowanych przykładów, wykorzystując do tego mocno kontrastujące kolory oraz technikę przezroczystości obrazu (Tab. 1).

Wyniki

Pierwsza seria danych składająca się ze zdjęć z NASA Earth Observatory [2] reprezentowała części lodowców oraz oderwane od nich kawałki lodu, które unosiły się na wodzie- przykład 1_1 i 1_2 (Tab. 1). Sklasyfikowane dryfujące kawałki lodu zostały zaznaczone półprzezroczystymi elementami

Tab. 1 Przykłady klasyfikacji w zależności od użytego progu binaryzacji

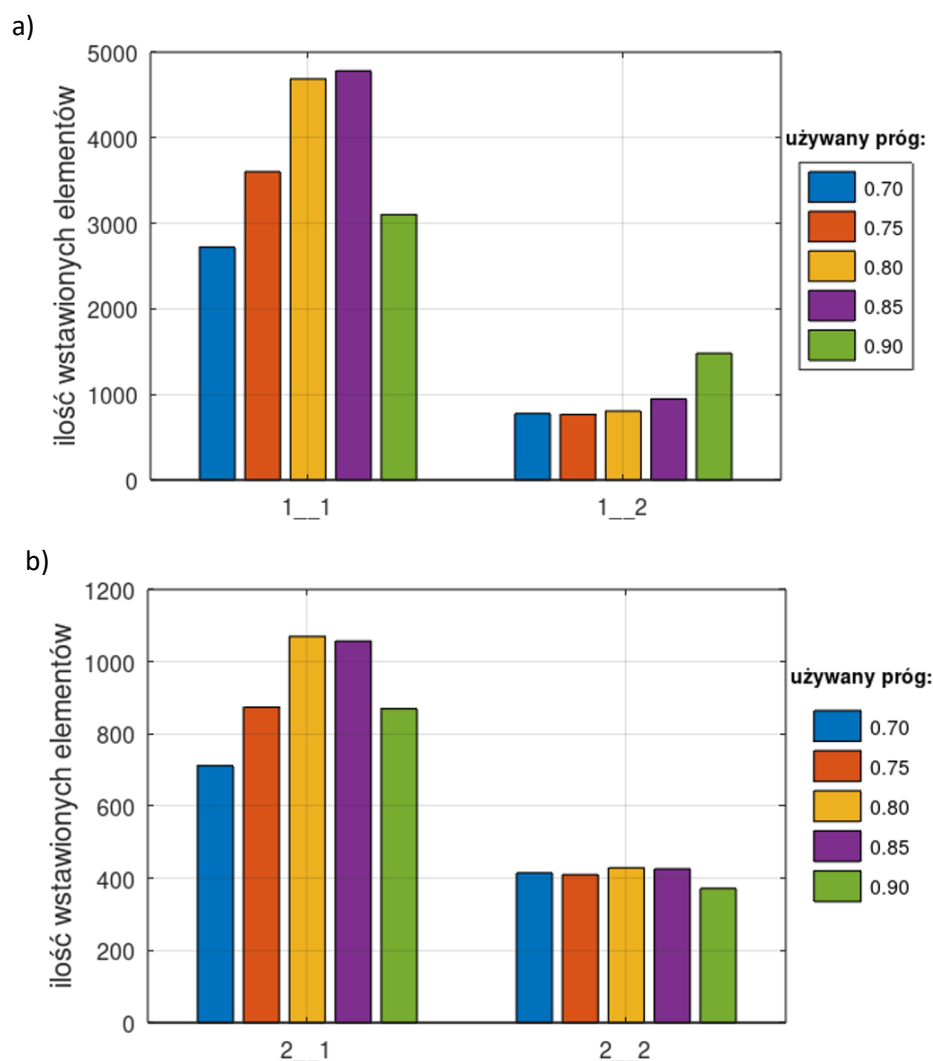
	1_1	1_2	2_1	2_2
oryginalny				
0.70				
0.75				
0.80				
0.85				
0.90				

w kształcie elips (Tab. 1). Dane użyte w drugiej części eksperymentu przedstawiały zdjęcia perspektywiczne czoła lodowców- przykład 2_1 i 2_2 (Tab. 1). W celu zobrazowania prawdopodobnych obszarów ich cielenia zostały nałożone półprzezroczyste, żółte elementy w kształcie czworokątów.

Ponadto tabela 1 przedstawia wyniki analiz przykładów 1_1 i 1_2 dla różnych progów binaryzacji. Można zauważyć, że mimo obecności drobnego lodu dookoła większych części, stabilniejsze fragmenty lodowca zostały odpowiednio sklasyfikowane.

Na przykładzie 2_2 widoczne są dodatkowo chmury w jasnym kolorze, które zostały błędnie odebrane jako naświetlone części lodowca, zakłámując tym samym dalszą parametryzację (Tab. 1; Ryc. 1). W analizie parametrycznej elementem zaburzającym analizę była woda, której jaśniejsze części również były identyfikowane (przy słabym doświadczeniu). Przez co zaburzały one parametryzację tak samo jak chmury.

Wyniki obrazujące ilość sklasyfikowanych elementów przy różnych progach binaryzacji zebrano na rycinie 1.



Ryc. 1. Liczba sklasyfikowanych elementów dla przykładów eksperymentu w zależności od progu binaryzacji

Zastosowanie analizy parametrycznej na przykładzie 1_1 (Ryc. 1a) wykazało, że mimo dużej ilości elementów, miały one niskie wartości dla długości obu osi opisujących ich kształt. Największy element wystąpił na zdjęciu 1_1 przy progu 0.70. Obejmował on prawie cały obszar, na którym drobny lód występował pomiędzy większymi kawałkami (Tab. 1). Sytuacja ta prawdopodobnie była spowodowana brakiem wystarczającego doświadczenia algorytmu (tj. zbyt mało analizowanych przykładów).

Przykład 2_1 cechował się dużą ilością zidentyfikowanych elementów, z czego największy z nich został zaobserwowany przy progu 0.80 (Ryc. 1b). W tych przykładach

wpływ na parametryzację mogły mieć chmury, które zostały wadliwie określone jako lód (Tab. 1). Jest to efekt niewystarczającego doświadczenia algorytmu (zbyt małego zbioru do analiz).

W dalszym etapie eksperymentu uwaga została poświęcona analizie entropii oraz jej zmianom w zależności od zmieniających się progów binaryzacji. Entropia jest miarą nieuporządkowania, im większe wartości przyjmują dane tym bardziej są one nieuporządkowane. Entropia może być informacją na temat skuteczności metody, wyższe wartości świadczą o większej złożoności obrazu, tj. słabszej parametryzacji [3].

Tab. 2 Wartości entropii kolorowych zdjęć oryginalnych przed analizą oraz zależność entropii od binaryzacji przykładów po analizie z obu części eksperymentu

	1_1	1_2	2_1	2_2
Zdjęcia oryginalne	6.55	5.62	7.61	7.09
0.70	3.81	2.79	3.68	4.61
0.75	4.01	2.79	3.67	5.27
0.80	4.21	2.79	3.67	5.28
0.85	4.32	2.74	3.67	5.27
0.90	4.58	2.83	3.68	5.27

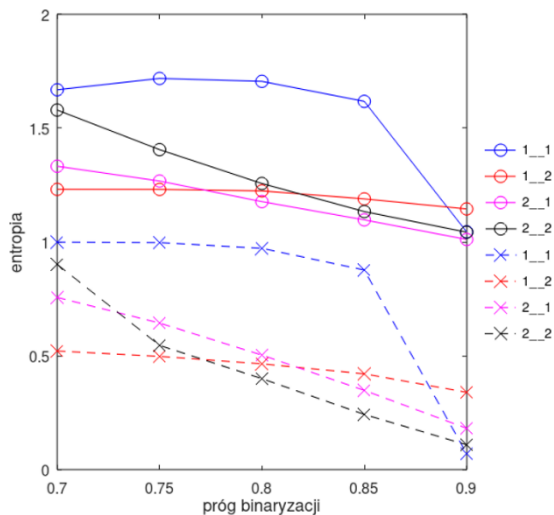
Rozpatrując zakres entropii od 0 do 8 bitów (Zbili, Rama 2021), można zauważyć, że wszystkie przykłady miały stosunkowo wysoką entropię (Tab. 2). Dane używane w drugiej części eksperymentu miały większą entropię.

Rozważając dane sparametryzowane algorytmem - entropia przyjmuje wartości z zakresu 2.74–5.28 bitów (Tab. 2), co jest znacznie niższe od entropii obliczonej dla kolorowych zdjęć oryginalnych przed analizą

(Tab. 2). Jest to spowodowane tym, że naniezione kształty przyczyniły się częściowo do rozmycia obrazu co zmniejsza entropię. Algorytm AI obniżył nieuporządkowanie obrazów - wykonał poprawnie parametryzację i okazał się skuteczny we wszystkich przykładach.

Porównując entropie oryginalnych zbinaryzowanych (Ryc. 2) danych z danymi zbinaryzowanymi po analizie (czarnobiałe zbinaryzowane z zaznaczonymi półprzezroczystymi żółtymi kształtami) (Ryc. 2) widać, że

wartości entropii są większe po niż przed analizą (Ryc. 2). Naniesione kształty, przez swoją różnorodność w porównaniu do 0-1 zbinaryzowanych danych podniosły ich entropię. W większości przypadków entropia ma mniejsze wartości wraz z rosnącym progiem binaryzacji (Ryc. 2).



Ryc. 2 Zależność entropii od binaryzacji dla zbinaryzowanych przykładów o - po analizie, x - przed analizą

Podsumowanie

Do celów eksperymentu wykorzystano AI, dzięki któremu proces binaryzacji został odpowiednio sparametryzowany. Wykazano, że próg binaryzacji ma zasadniczy wpływ na wynik. Na analizowanych przykładach większe elementy lodu zostały odpowiednio oddzielone od skruszonych, drobnych kawałków i od śniegu. Można zauważyć, że wraz ze wzrostem progu wstawiane elementy są mniejsze – wyznacza to kierunek nauczania maszynowego systemu. W części przykładów przeszkodą okazała się obecność chmur. Z tego powodu dokładniejsze reprezentacje są widoczne przy wyższych progach. Na analizowanych przykładach obszary oświetlone zostały bardzo dobrze oznaczone przez AI w określonych progach.

Metoda przedstawiona w pracy charakteryzuje się dużym potencjałem do wykorzystania przy pracy z danymi niestabilnymi (trudnymi w parametryzacji tradycyjnymi metodami). Natomiast w celu weryfikacji możliwych obszarów cielenia się lodowców może być wymagana większa ilość danych z racji chociażby różnych obszarów oświetlenia w zależności od pozycji słońca i innych czynników od tego niezależnych wpływających na destabilizację lodowców. Na analizowanych zdjęciach satelitarnych metoda spełniała swoje założenia. Działała ona szczególnie dobrze w przypadku wykrywania większych części lodu obszarów pokrytych drobnym lodem i śniegiem co może być przydatne w pracy z podobnymi, trudniejszymi w wizualnej interpretacji, danymi.

Praca obejmowała analizę serii danych oceanograficznych przedstawiających lodowce i lód na tle wody. Na podstawie przytoczonego eksperymentu można stwierdzić, że wykorzystane elementy algorytmu nauczania maszynowego działają przy parametryzacji niestabilnych serii danych. Zasadniczą rolę odegrało wyznaczenie progu binaryzacji - to zadanie powierzono AI. Generalnie im więcej danych dostarczymy tym algorytm AI lepiej sobie radzi. Oprócz wizualnej reprezentacji zaznaczonych elementów uzyskano także parametry numeryczne przykładów np: długości osi głównej i pomocniczej elementów, ilości wstawionych elementów oraz ich orientacji. W oparciu o te parametry można dokonywać dalszej analizy np.: mając skalę można prowadzić analizy wielkości wstawionych elementów, które mogą dać informacje o rozmiarach poszczególnych elementów dryfującego lodu lub rozmiarach powierzchni oświetlonych lodowców. Zasadniczy dowód skuteczności metody uzyskano poprzez porównanie entropii serii przykładów eksperymentalnych.

Literatura:

- Ahmad H. (2019). Machine learning applications in oceanography. *Aquatic Research*, nr 2(3), s. 161-169.
- Alpaydin E. (2021). *Machine learning*. Cambridge: Wydawnictwo Mit Press.
- Duvillard P.A., Ravanel L., Deline P. (2015). Risk assessment of infrastructure destabilisation due to global warming in the high French Alps. *Journal of Alpine Research/ Revue de géographie alpine*, nr 103(2), s. brak.
- Ekasari S., Judijanto L., Vandika, A. Y. (2024). The role of artificial intelligence and machine learning in precision targeting: revolutionizing marketing. *Jurnal Ekonomi dan Bisnis (JEBI)*, nr 2(3), s. 347-361.
- Faillietaz J., Funk M., Vincent C. (2015). Avalanching glacier instabilities: Review on processes and early warning perspectives. *Reviews of Geophysics*, nr 53(2), s. 203-224.
- Haeberli W., Schaub Y., Huggel C. (2017). Increasing risks related to landslides from degrading permafrost into new lakes in deglaciating mountain ranges. *Geomorphology*, nr 293, s. 405-417.
- Marcer M., Serrano C., Brenning A., Bodin X., Goetz J., Schoeneich P. (2019). Evaluating the destabilization susceptibility of active rock glaciers in the French Alps. *The Cryosphere*, nr 13(1), s. 141-155.
- Molvaer, R. K. (1991). Environmentally induced conflicts? A discussion based on studies from the Horn of Africa. *Bulletin of Peace Proposals*, nr 22(2), s. 175-188.
- Paszkuta M. (2023). *2AI-Artificial Intelligence as a key component in inspiring education through Authentic Inquiry*. Prezentowane na: International Conference "STEAM & AI in Education". XII International Scientific Seminar "Science - Society - Didactics"
- Paszkuta M., Szymczak E. (2024). *Using AI to develop academic skills in practical university courses*. Prezentowane na: International Conference "STEAM & AI in Education". XIII International Scientific Seminar "Science - Society - Didactics"
- Ponomarenko I.V., Pavlenko V.M., Morhulets O.B., Ponomarenko D.V., Ukhna N.M. (2024). Application of artificial intelligence in digital marketing. W: *CEUR Workshop Proceedings*, nr 3662, s. 155-166.
- Stocker A.N., Renner A.H.H., Knol-Kauffman M. (2020). Sea ice variability and maritime activity around Svalbard in the period 2012–2019. *Scientific Reports*, nr 10, s. brak.
- Woodman R. J., Mangoni A. A. (2023). A comprehensive review of machine learning algorithms and their application in geriatric medicine: present and future. *Aging Clinical and Experimental Research*, nr 35(11), s. 2363-2397.
- Zakhvatkina N., Smirnov V., Bychkova I. (2019). Satellite SAR data-based sea ice classification: An overview. *Geosciences*, nr 9(4), s. 152.
- Zbili M., Rama S. (2021). A quick and easy way to estimate entropy and mutual information for neuroscience. *Frontiers in Neuroinformatics*, nr 15(596443).

Źródła internetowe:

- 1_1: NASA Earth Observatory <https://earthobservatory.nasa.gov/topic/snow-and-ice> [dostęp 02.11.2023]
- 1_2: NASA Earth Observatory <https://earthobservatory.nasa.gov/topic/snow-and-ice> [dostęp 02.11.2023]
- 2_1: Kukulya A. <https://www.whoi.edu/know-your-ocean/ocean-topics/how-the-ocean-works/frozen-ocean/glaciers-ice-sheets/>
- 2_2: Mommaerts R. <https://timesofindia.indiatimes.com/travel/things-to-do/hubbard-glacier/articleshow/35384197.cms>
- [1] - <https://www.sap.com/poland/products/artificial-intelligence/what-is-machine-learning.html> [dostęp 31.08.2023]
- [2] - <https://earthobservatory.nasa.gov/topic/snow-and-ice> [dostęp 02.11.2023]

[3] - <https://unimatrixz.com/blog/latent-space-image-quality-with-entropy/> [dostęp 06.02.2024]

Notka o autorce: Aleksandra Kaleta ukończyła studia licencyjne na kierunku oceanografia. Swoje zainteresowania związane z geofizyką zdecydowała się rozwijać podczas zajęć tutoringu akademickiego na WOiG UG. Realizując pracę dyplomową skupiła swoją uwagę na nowoczesnych metodach analizy lodowców, na podstawie danych uzyskiwanych z ekspedycji pod koło polarne. Bardzo szybko jej ciekawość zwróciła się ku szerszemu wykorzystaniu nowoczesnych technik nauczania maszynowego w analizie powierzchni morza na podstawie danych satelitarnych.