



Kierunek studiów: Geoinformacja

Nr albumu: 461791

Filip Ratajszczak

**Wykrywanie farm fotowoltaicznych na podstawie
danych teledetekcyjnych**

My title

Praca inżynierska napisana
w Instytucie Geoekologii i Geoinformacji
pod kierunkiem
dr. hab. Jakuba Nowosada

Poznań, 2024

Streszczenie

Abstrakt

Streszczenie powinno przedstawiać skrótowo główny problem pracy i jego rozwiązanie. Możliwa struktura streszczenia to: (1) 1-3 zdania wstępu do problemu (czym się zajmujemy, dlaczego jest to ważne, jakie są problemy/luki do wypełnienia), (2) 1 zdanie opisujące cel pracy, (3) 1-3 zdania przedstawiające użyte materiały (dane) i metody (techniki, narzędzia), (4) 1-3 zdania obrazujące główne wyniki pracy, (5) 1-2 zdania podsumowujące; możliwe jest też określenie dalszych kroków/planów.

Słowa kluczowe: (4-6 słów /zwrotów opisujących treść pracy, które nie wystąpiły w tytule)

Abstract

The abstract must be consistent with the above text.

Keywords: (as stated before)

Spis treści

Streszczenie	3
1 Wprowadzenie	5
2 Przegląd literatury	7
2.1 Podrozdział	8
3 Materiały	9
3.1 Zdjęcia satelitarne	9
3.2 Ortofotomapa i mozaiki zdjęć satelitarnych	14
4 Metody	17
4.1 Przygotowanie danych	17
4.2 Próbki treningowe i testowe	19
4.3 Uczenie maszynowe	19
4.4 Oprogramowanie	21
5 Wyniki	25
6 Podsumowanie	27

Rozdział 1

Wprowadzenie

Wprowadzenie powinno mieć charakter opisu od ogółu do szczegółu (np. trzy-pięć paragrafów). Pierwszy paragraf powinien być najbardziej ogólny, a kolejne powinny przybliżać czytelnika do problemu. Przedostatni paragraf powinien określić jaki jest problem (są problemy), który praca ma rozwiązać i dlaczego jest to (są one) ważne.

Wprowadzenie powinno być zakończone stwierdzeniem celu pracy. Dodatkowo tutaj może znaleźć się również krótki opis co zostało zrealizowane w pracy.

Rozdział 2

Przegląd literatury

Ten rozdział zawiera wyjaśnienie kontekstu pracy.

Pisząc ten rozdział proszę pomyśleć o osobach, które zupełnie nie znają opisywanej tematyki. Należy tutaj krok po kroku wyjaśnić podstawowe koncepcje, istotność problemu, wyniki poprzednich podobnych badań, itd. Ten rozdział obejmuje tylko kwestie, które już zostały wykonane przez inne osoby - nowe wyniki mają swoje miejsce w rozdziale 5.

Każda kwestia opisana w tym rozdziale powinna być cytowana. Dodatkowo cytowania odbywa się poprzez uzupełnienie pliku `thesis.bib` zapisem w formacie BibTeX, a następnie dodanie nazwy referencji poprzedzonej znakiem @. Przykładowo, zacytowanie książki *Geocomputation with R* odbywa się poprzez (`lovelace_geocomputation_2019`).

W przypadku, gdy cytowanie zostało poprawnie wpisane oraz istnieje w pliku `thesis.bib` to bibliografia powinna się automatycznie wygenerować na końcu pracy.

W przypadku, gdy praca dyplomowa opisuje konkretny obszar to można po tym rozdziale stworzyć kolejny rozdział opisujący "obszar badań".

Ten i kolejne rozdziały mogą mieć także podrozdziały. Tworzenie podrozdziałów polega na stworzeniu nowej linii rozpoczynającej się od znaków `##` a następnie tytułu podrozdziału. Dodatkowo w postaci `{#sec-}` można dodać skrót nazwy rozdziału/podrozdziału umożliwiając odnoszenie się do niego używając operatora `[-@sec]`.

2.1 Podrozdział

Przykładowo, “te kwestie zostały opisane w podrozdziale 2.1”. Zwróć uwagę, że w ten sposób automatycznie tworzony jest odnośnik w pliku PDF.

Rozdział 3

Materiały

3.1 Zdjęcia satelitarne

3.1.1 Sentinel-1

Misja Sentinel-1 stanowi Europejskie Obserwatorium Radarowe działające w ramach programu Copernicus, wspólnej inicjatywy Komisji Europejskiej i Europejskiej Agencji Kosmicznej (ESA, [n.d.\[e\]](#)). Celem misji jest dostarczanie danych radarowych obejmujących globalną powierzchnię Ziemi, w tym lądów, europejskich stref przybrzeżnych, tras żeglugowych, stref lodu morskiego, mórz i oceanów (Hejmanowska et al., [2020](#); ESA, [n.d.\[d\]](#)). Misja odpowiada na potrzeby monitorowania obszarów morskich i lądowych, w tym przemieszczania kry lodowej, transportu morskiego, deformacji i przemieszczeń terenu, a także obserwacji zmian klimatu i klęsk żywiołowych (Hejmanowska et al., [2020](#); ESA, [n.d.\[d\]](#)).

Sentinel-1 jest wyposażony w radar z syntetyczną aperturą (ang. *Synthetic Aperture Radar*, SAR), pracujący w paśmie C na częstotliwości 5,405 GHz, co odpowiada długości fali 5,6 cm (Braun, [2020](#); ESA, [n.d.\[c\]](#)). Radar działa głównie w trybie podwójnej polaryzacji, emitując fale pionowe i mierząc zarówno fale pionowe (ang. *vertical*), jak i poziome (ang. *horizontal*) po ich powrocie do czujnika, dzięki czemu otrzymujemy dane o intensywności rozproszenia wstecznego VV i VH (Braun, [2020](#)). Radar z syntetyczną aperturą (SAR)

umożliwia prowadzenie obserwacji zarówno w nocy, jak i w dzień, niezależnie od warunków pogodowych, przez co czas między rejestracją obrazów dla pojedynczego satelity Sentinel-1 wynosi 12 dni nad równikiem, a dla konstelacji dwóch satelitów skraca się do od dwóch do sześciu dni, zależnie od szerokości geograficznej (Hejmanowska et al., 2020; ESA, n.d.[e]).

Dane radarowe Sentinel-1 są dostępne na trzech poziomach przetworzenia. Produkty poziomu 1 (Single Look Complex (SLC) i Ground Range Detected (GRD)), są przeznaczone dla użytkowników końcowych i nadają się, w zależności od produktu do monitorowania Ziemi, klasyfikacji pokrycia terenu czy aplikacji interferometrycznych (Hejmanowska et al., 2020). SLC to obrazy SAR w geometrii ukośnej, zachowujące informację fazową sygnału (Hejmanowska et al., 2020). Produkty GRD są rezultatem przepróbkowania obrazów SLC do jednolitej rozdzielczości przestrzennej (10×10 m) i rzutowania na powierzchnię elipsoidy odniesienia (Hejmanowska et al., 2020). GRD nie zawiera cech ortofotomapy i wymaga dodatkowego przepróbkowania z użyciem numerycznego modelu terenu przed zastosowaniem w systemach GIS (Hejmanowska et al., 2020). Przy konwersji SLC do GRD tracimy informację fazową sygnału, co sprawia, że produkty GRD nie są odpowiednie do interferometrii radarowej (ESA, n.d.[b]).

Dane Sentinel-1 rejestrowane są w czterech trybach: Interferometric Wide Swath (IW) – podstawowy tryb obrazowania dla obszarów lądowych, Stripmap (SM) – używany do obrazowania małych wysp i na potrzeby zarządzania kryzysowego, Extra-Wide Swath (EW) – tryb do monitorowania stref polarnych i niektórych obszarów morskich, Wave (WV) – tryb obrazowania oceanów (Hejmanowska et al., 2020; ESA, n.d.[c],[a]).

Dane wykorzystane do analizy pochodzą z produktów Ground Range Detected (GRD) zarejestrowanych w trybie Interferometric Wide Swath (IW) przez satelitę Sentinel-1B w dniu 8 maja 2023 roku. Zestaw danych został utworzony poprzez połączenie dwóch sąsiednich produktów Sentinel-1 GRD, a następnie ograniczenie obszaru analizy do kafla Sentinel-2 o oznaczeniu 33UWV. W analizie uwzględniono dwie polaryzacje: VV i VH.

3.1.2 Sentinel-2

Misja Sentinel-2 stanowi inicjatywę Komisji Europejskiej, która jest operacyjnie prowadzona przez Europejską Agencję Kosmiczną (ang. *European Space Agency*, ESA) w ramach programu Copernicus. Celem tej misji jest dostarczanie obrazów satelitarnych, obejmujących trzynaście zakresów spektralnych o różnych rozdzielczościach przestrzennych: 10, 20 lub 60 metrów, zależnie od rejestrowanego kanału. Rozdzielczość czasowa misji Sentinel-2 wynosi pięć dni nad równikiem i zwiększa się wraz ze wzrostem szerokości geograficznej, osiągając dwa dni na średnich szerokościach geograficznych (Hejmanowska et al., 2020).

Dane pozyskiwane przez satelity Sentinel-2 są dostępne na różnych poziomach przetworzenia, lecz najczęściej używane przy tworzeniu map pokrycia terenu i użytkowania ziemi (ang. *Land Use/Land Cover*, LULC) są produkty 1C (współczynnik odbicia na poziomie górnej części atmosfery; ang. *Top-of-Atmospheric reflectance*, TOA) oraz 2A (współczynnik odbicia na powierzchni Ziemi; ang. *Bottom-of-Atmospheric reflectance*, BOA) (Phiri et al., 2020).

Produkty poziomu 1C to dane poddane korekcjom radiometrycznym i geometrycznym, prezentowane jako sceny o powierzchni 100 km² (100 x 100 km) w projekcji UTM/WGS84 (ESA, 2015). Skuteczne wykorzystanie tych danych w zastosowaniach związanych z terenami lądowymi wymaga precyzyjnej korekcji zdjęć satelitarnych pod kątem efektów atmosferycznych (Main-Knorn et al., 2017). Produkty poziomu 2A powstają poprzez zastosowanie dodatkowej korekcji atmosferycznej dla danych poziomu 1C za pomocą procesora korekcji atmosferycznej Sen2Cor (Main-Knorn et al., 2017).

Dane wykorzystane w analizie pochodzą z dnia 8 maja 2023 roku i zostały dostarczone przez satelitę Sentinel-2B. Obszar analizy obejmuje kafel (ang. *tile*) o oznaczeniu 33UWV, dla którego współczynnik zachmurzenia w tym dniu wynosił 0,7%. Użyte zostały dane na poziomie przetworzenia L2A. Z dostępnych kanałów spektralnych (Tabela 3.1) wykorzystano 10 zakresów, ponieważ pasma rejestrowane w rozdzielczości 60 metrów są przeznaczone głównie do korekcji atmosferycznych i detekcji chmur. Kanał 1 (443

Tabela 3.1: Kanały spektralne satelitów Sentinel-2

Kanał	Nazwa kanału	Centralna długość fali [nm]	Zakres spektralny [nm]	Rozdzielczość przestrzen- na [m]
B01	Coastal Aerosol	443	433–453	60
B02	Blue	490	458–523	10
B03	Green	560	543–578	10
B04	Red	665	650–680	10
B05	Vegetation RedEdge	705	698–713	20
B06	Vegetation RedEdge	740	733–748	20
B07	Vegetation RedEdge	783	773–793	20
B08	NIR	842	785–900	10
B8A	NIR	865	855–875	20
B09	Water Vapour	945	935–955	60
B10	Cirrus	1375	1360–1390	60
B11	SWIR	1610	1565–1655	20
B12	SWIR	2190	2100–2280	20

nm) służy do korekcji wpływu aerozoli, kanał 9 (940 nm) do korekcji wpływu pary wodnej, a kanał 10 (1375 nm) do wykrywania chmur typu cirrus (Drusch et al., 2012).

3.1.3 Indeksy spektralne

Oprócz surowych współczynników odbicia, w niektórych wariantach predykcji zastosowano również wskaźniki spektralne (ang. *spectral indices*), które były wykorzystywane w poprzednich badaniach przez Zhang et al. (2021), Plakman et al. (2022), Wang et al. (2022) i innych, takie jak:

- znormalizowany wskaźnik roślinności (ang. *Normalized Difference Vegetation Index*, NDVI) (Tucker, 1979), monitorujący zawartość biomasy i kondycję roślinności na danym obszarze

$$NDVI = \frac{NIR - Red}{NIR + Red}$$

, gdzie *NIR* – refleksja w kanale bliskiej podczerwieni, *Red* – refleksja w kanale czerwonym

- znormalizowany indeks zabudowy (ang. *Normalized Difference Built-up Index*, NDBI) (Zha et al., 2003), przeznaczony do kartowania obszarów zabudowanych

$$NDBI = \frac{SWIR1 - NIR}{SWIR1 + NIR}$$

, *SWIR1* – refleksyjność w kanale średniej podczerwieni, *NIR* – refleksyjność w kanale bliskiej podczerwieni

- znormalizowany indeks wody (ang. *Modified Normalized Difference Water Index*, mNDWI) (Xu, 2006), który skutecznie identyfikuje obszary wodne na zdjęciach satelitarnych, posiadając możliwości tłumienia zakłóceń spowodowanych przez zabudowę, roślinność i gleby

$$mNDWI = \frac{Green + SWIR1}{Green - SWIR1}$$

, gdzie *Green* – refleksyjność w kanale zielonym, *SWIR1* – refleksyjność w kanale średniej podczerwieni

3.1.4 Tekstury obrazu

Tekstura stanowi istotną cechę wykorzystywaną do identyfikacji obiektów i obszarów zainteresowania na obrazie (Haralick et al., 1973) i odgrywa ona dużą rolę w interpretacji wizualnej zdjęć lotniczych i satelitarnych (Lewiński, 2012). Gdy różnice widmowe pomiędzy klasami są niewielkie, tekstura umożliwia rozróżnienie odmiennych typów obiektów na podstawie ich organizacji w terenie, często kontrastując przestrzenie naturalne z antropogenicznymi (GRASS, 2023). W zależności od zastosowanej funkcji wybrane cechy obrazu zostają uwidocznione w porównaniu z jego obrazem wejściowym (Lewiński, 2012). Informacja o teksturze może stanowić dodatkową, przydatną zmienną wejściową w procesach klasyfikacji lub segmentacji obrazu (Gong et al., 1992; Mumby et al., 2002). Tekstura obejmuje różnice poziomów szarości (kontrast), obecność lub brak kierunkowości, regularne wzory i zdefiniowany obszar, na którym występują zmiany, określony przez rozmiar okna (Hall-Beyer, 2017; GRASS, 2023). Można ją opisać za pomocą tonu (intensywność poziomu szarości) i struktury (relacje przestrzenne) (GRASS, 2023).

Model tekstury oparty na macierzy współwystępowania poziomów szarości (ang. *Gray Level Co-Occurrence Matrix*, GLCM), zaproponowany przez Haralick et al. (1973), jest często używany do obliczania tekstur obrazu. Ta metoda polega na tworzeniu macierzy opisującej częstotliwość występowania par wartości w określonym fragmencie obrazu, uwzględniając określone sąsiedztwo, kierunki i przerwy między pikselami (Kupidura, 2019).

Przydatność i wykorzystanie tekstury w dużym stopniu zależy od rozdzielczości zdjęć satelitarnych i wielkości zjawiska, które ukształtowało teksturę (GRASS, 2023). Badanie, które przeprowadził (Zhang et al., 2021) dotyczące wykorzystania filtracji teksturalnych w identyfikacji elektrowni fotowoltaicznych z użyciem Random Forest i danych z Landsata-8 wykazało pozytywny wpływ tekstur na skuteczność modelu. Według wyników badania najlepiej dopasowany model wykorzystywał tekstury GLCM o sąsiedztwie 30 pikseli (co odpowiada wymiarom ruchomego okna o wymiarach 1800 m na 1800 m), natomiast tekstura o rozmiarze jednego sąsiada ma niewielki wpływ na poprawę dokładności modelu (Zhang et al., 2021).

Obliczanie tekstur obrazu może być czasochłonnym procesem, dlatego w pracy skorzystano jedynie z tekstury średniej sumy (ang. *Sum Average*, SA lub SAVG), wskazanej przez Wang et al. (2022) jako teksturę niosącą najwięcej informacji w kontekście detekcji farm fotowoltaicznych na podstawie danych Sentinel-1, Sentinel-2 i algorytmu Random Forest. W celu skrócenia czasu obliczeń, w pracy zdecydowano się na zastosowanie ruchomego okna o sąsiedztwie 9 komórek.

3.2 Ortofotomapa i mozaiki zdjęć satelitarnych

Do lokalizacji oraz wektoryzacji istniejących farm fotowoltaicznych wykorzystano ortofotomapę udostępnianą przez Główny Urząd Geodezji i Kartografii oraz mozaiki zdjęć satelitarnych dostarczane przez podmioty komercyjne. W teledetekcji jednym z zastosowań mozaik obrazów satelitarnych jest tworzenie zestawów danych referencyjnych poprzez interpretację wizualną, np. w celu walidacji wyników klasyfikacji produktów pokrycia terenu (Lesiv et al., 2018). Do stworzenia zbioru danych testowych i treningowych

wykorzystano ortomozaiki Google Satellite, Bing Aerial, Mapbox Satellite oraz Planet Basemaps, udostępniane w formie usług sieciowych (WMS, WMTS, XYZ Tiles). Ortomozaiki te są tworzone na podstawie komercyjnych zdjęć satelitarnych wykonywanych przez podmioty takie jak Maxar Technologies, Airbus czy Planet Labs.

Ortofotomapa udostępniana przez GUGiK charakteryzuje się rozdzielczością przestrzenną 25 cm, a rozdzielczość przestrzenna wykorzystanych mozaik obrazów satelitarnych (poza Planet Basemaps) jest wyższa niż 1 m, np. dla mozaiki Bing Aerial udostępnianej przez firmę Microsoft wynosi ona 30-60 cm (GUGiK, [n.d.](#); Microsoft, [2020](#)). Często jednak nie jest możliwe ustalenie dat wykonania zdjęć satelitarnych, które posłużyły do stworzenia konkretnej mozaiki zobrazowań satelitarnych (Lesiv et al., [2018](#)). Mozaika tworzona przez Planet na podstawie zdjęć satelitarnych wykonywanych przez konstelację satelitów PlanetScope charakteryzuje się rozdzielczością przestrzenną 4,77 m na równiku, jednak w porównaniu do pozostałych wymienionych produktów jest tworzona z miesięczną oraz kwartalną częstotliwością (Planet Labs Inc., [2019](#)). Pozwala to, pomimo niższej rozdzielczości przestrzennej na stworzenie zbioru danych testowych i treningowych na konkretny okres czasu. Mozaiki Planet Basemaps są tworzone na podstawie najlepszych obrazów z katalogu Planet w określonym przedziale czasowym. Wybierając najlepsze zdjęcia, Planet jest w stanie tworzyć wysokorodzielcze mozaiki, które są dokładne radiometrycznie i przestrzennie, a także charakteryzują się zminimalizowanym wpływem czynników atmosferycznych (Planet Labs Inc., [2019](#)).

?Rycina przedstawiająca jedną farmę na ortofotomapie, mozaikach satelitarnych i scenie Sentinel-2 (kompozycja RGB)

Rozdział 4

Metody

4.1 Przygotowanie danych

4.1.1 Sentinel-1

Korzystanie z danych radarowych wymaga wcześniejszego przygotowania danych poprzez proces kalibracji, aby zapewnić poprawne wyniki analizy. Procesy te mogą różnić się w zależności od konkretnego zastosowania, mając na celu dostosowanie danych do specyficznych potrzeb. W pracy wykorzystano schemat przetwarzania danych Sentinel-1 GRD, który zaproponował Filipponi ([2019](#)), obejmujący:

1. aktualizację danych o położeniu satelity w momencie zobrazowania poprzez pobranie i zastosowanie pliku orbity;
2. korekcję szumów termicznych;
3. korekcję szumów na granicach obrazów;
4. obliczenie współczynnika rozproszenia wstecznego (ang. *backscatter coefficient*) σ_0 za pomocą kalibracji radiometrycznej;
5. korekcję topograficzną (ortorektyfikacja za pomocą Copernicus 30 m Global DEM);
6. konwersję współczynnika rozproszenia wstecznego na dB za pomocą transformacji logarytmicznej

rycina - Sentinel-1 GRD Preprocessing Workflow

Wstępne Przetwarzanie danych dla obu wykorzystywanych polaryzacji (VV i VH) zostało wykonane przy użyciu zestawu narzędzi ESA Sentinel-1 Toolbox (S1TBX) (ESA, 2023a) w oprogramowaniu SNAP (ESA, 2023d) przy pomocy narzędzie do przetwarzania grafów (ang. *Graph Processing Tool*, GPT) . Kolejne etapy przygotowania danych zostały zrealizowane przy wykorzystaniu języka R (R Core Team (2023)) oraz pakietu *terra* (Hijmans, 2023). Obszar analizy, będący kaflem Sentinel-2 o oznaczeniu 33UWV, znajduje się na granicy dwóch sąsiednich produktów Sentinel-1 GRD. Na potrzeby dalszego przetwarzania, sąsiadujące produkty zostały połączone i odpowiednio ograniczone do obszaru zainteresowania. Na granicy sąsiednich produktów Sentinel-1 GRD występowała przestrzeń bez danych, co wymagało wypełnienia tego obszaru danymi przy użyciu funkcji *focal* z pakietu *terra*.

rycina - mapa produkty S1 GRD (granice) vs. granice kafla 33UWV Sentinel-2

4.1.2 Sentinel-2

Przetwarzanie danych Sentinel-2 polegało na sprowadzeniu kanałów o rozdzielczości 20 m do rozdzielczości i siatki kanałów w rozdzielczości 10 m. Przeprobkowanie (ang. *resampling*) zostało przeprowadzone przy wykorzystaniu języka R (R Core Team (2023)) oraz funkcji *resample* z pakietu *terra* (Hijmans, 2023), wykorzystując interpolację dwuliniową (ang. *bilinear interpolation*).

4.1.3 Łączenie danych

Rozdzielczość przestrzenna danych Sentinel-1 GRD, podobnie jak danych Sentinel-2 przygotowanych w sposób przedstawiony w podrozdziale 4.1.2 wynosi 10 m. Stworzenie spójnych wielokanałowych rastrów wymaga sprowadzenia wszystkich zestawów danych do wspólnej rozdzielczości i siatki. Dane Sentinel-1 GRD zostały przetransformowane do siatki danych Sentinel-2. Przeprobkowane do wspólnej rozdzielczości oraz siatki dane zostały następnie użyte do obliczeń filtracji teksturalnych oraz wskaźników teledetekcyjnych. Po uzyskaniu produktów pochodnych, w zależności od wariantu, dane zostały

scalone w kilka wielokanałowych rastrów, które posłużyły do wyodrębnienia zestawu danych treningowych oraz do przeprowadzenia predykcji.

tabela - warianty połączenia danych - np. kanały S2, kanały S2 + indeksy spektralne, kanały S2 + polaryzacje S1 etc.

rycina - schemat przygotowania danych

4.2 Próbkowanie treningowe i testowe

Opis wektoryzacji farm fotowoltaicznych na podstawie ortofotomapy i mozaik satelitar-nych. Opis procesu pozyskiwania próbek treningowych, uwzględniając liczbę próbek pozytywnych i negatywnych, ewentualne dobieranie negatywnych próbek z jakiejś konkretnej kategorii pokrycia terenu - do napisania po ostatecznym generowa-niu/losowaniu próbek

4.3 Uczenie maszynowe

Klasyfikacja obrazów w teledetekcji polega na grupowaniu komórek w niewielkie zestawy klas, aby komórki w tych samych klasach miały podobne właściwości (Ismail et al., 2009). Istnieje wiele różnych metod klasyfikacji danych teledetekcyjnych. Stosunkowo nowymi podejściami wykorzystywanymi w tym kontekście są metody oparte na sztucznej inteligencji, takie jak uczenie maszynowe (ang. *Machine Learning*, ML) lub uczenie głębokie (ang. *Deep Learning*, DL) (Hejmanowska et al., 2020).

Uczenie maszynowe stanowi obszar sztucznej inteligencji, koncentrujący się na opracowywaniu algorytmów i modeli statystycznych zapewniających systemom komputerowym możliwość automatycznego uczenia się z danych i wykonywania określonych zadań bez konieczności bezpośredniego programowania. W przypadku skomplikowanych i złożonych zestawów danych nie jesteśmy w stanie odpowiednio ich zinterpretować oraz wydobyć poprawnych informacji po wizualnym przejrzaniu danych (Mahesh, 2019). Uczenie maszynowe jest wykorzystywane do uczenia maszyn efektywnego przetwarzania danych (Sindayigaya et al., 2022). Algorytmy uczenia maszynowego można podzielić na cztery główne podejścia: uczenie nienadzorowane (ang. *unsupervised learning*),

uczenie nadzorowane (ang. *supervised learning*), uczenie częściowo nadzorowane (ang. *semi-supervised learning*) oraz uczenie przez wzmacnianie (uczenie posiłkowane, ang. *reinforcement learning*) (Sarker, 2021).

W ciągu ostatnich dwudziestu lat zaproponowano kilka różnych algorytmów uczenia maszynowego do klasyfikacji obrazów satelitarnych (Sheykhmousa et al., 2020), zazwyczaj wykorzystujące techniki klasyfikacji bez nadzoru i klasyfikacji nadzorowanej (Ismail et al., 2009).

Uczenie nienadzorowane analizuje nieoznakowane zbiory danych bez konieczności ingerencji człowieka. Uczenie bez nadzoru jest powszechnie stosowane do eksploracji danych, ekstrakcji cech generatywnych, identyfikacji istotnych trendów i struktur oraz grupowania wyników. Ta technika uczenia maszynowego jest najczęściej używana do grupowania (klastowania), redukcji wielowymiarowości (redukcji cech) oraz identyfikacji skojarzeń i relacji (Sarker, 2021).

Nadzorowane algorytmy uczenia maszynowego wykorzystują oznaczone dane treningowe do znajdowania powiązań pomiędzy różnymi zmiennymi. Proces uczenia nadzorowanego zachodzi, gdy określone cele mają zostać osiągnięte na podstawie konkretnego zestawu danych wejściowych (treningowych). Dwa główne typy uczenia nadzorowanego to klasyfikacja, która separuje dane, oraz regresja, która dopasowuje dane (Sarker, 2021).

W badaniu do klasyfikacji wykorzystano nadzorowaną metodę lasów losowych (ang. *Random Forest*, RF) (Breiman, 2001).

4.3.1 Metoda lasów losowych

Random Forest stał się jednym z najpopularniejszych klasyfikatorów uczenia maszynowego wykorzystywanych przez społeczność teledetekcyjną ze względu na dokładność jego klasyfikacji oraz wysoką wydajność (Belgiu et al., 2016; Sheykhmousa et al., 2020). Metoda lasów losowych charakteryzuje się odpornością na szumy (ang. *noise*) i przeuczenie (ang. *overfitting*), ponieważ nie bazuje na ważeniu (Gislason et al., 2006).

Algorytm Random Forest, będący rozwinięciem koncepcji drzew decyzyjnych, operuje na zasadzie ensemble learning, czyli łączenia wielu słabszych modeli (indywidualnych drzew decyzyjnych) w jeden silniejszy model (Maxwell et al., 2018; Sekulić et al., 2020). Procedura generuje liczne drzewa decyzyjne, opierając się na losowo wybranym zestawie danych ze zbioru danych uczących oraz losowo wyselekcjonowanych zmiennych klasyfikacyjnych (Breiman, 2001). Pojedyncze drzewo korzysta ze zredukowanej liczby danych treningowych i zmiennych, co sprawia, że drzewa różnią się od siebie i są mniej precyzyjne, ale jednocześnie są też mniej skorelowane, przez co model złożony z wielu drzew będzie bardziej niezawodny (Sekulić et al., 2020). W fazie predykcji każde z drzew w lesie dokonuje prognozy, a ostateczna decyzja jest formułowana na podstawie głosowania większościowego. W przypadku klasyfikacji, klasa wybierana jest na podstawie największej liczby głosów. (Malinowski et al., 2020).

4.4 Oprogramowanie

4.4.1 QGIS

QGIS, dawniej Quantum GIS (QGIS Development Team, 2023), to wieloplatformowe i wolne oprogramowanie o otwartym kodzie źródłowym przeznaczone do przetwarzania danych przestrzennych, rozwijane przez QGIS Development Team od 2002 roku (Hejmanowska et al., 2020; Flenniken et al., 2020). Algorytmy przetwarzania danych przestrzennych zebrane w oprogramowaniu QGIS umożliwiają manipulację danymi rastrowymi oraz wektorowymi, a także prowadzenie analiz i wizualizację wyników (Hejmanowska et al., 2020). Oprogramowanie QGIS oferuje również możliwość korzystania z wielu zewnętrznych programów, tzw. wtyczek (ang. *plug-in*) rozszerzających jego funkcjonalność (Hejmanowska et al., 2020). W repozytorium wtyczek znaleźć można narzędzia do zarządzania danymi, przetwarzania obrazów, wizualizacji, czy wykonania dodatkowych zadań, takich jak np. nadawanie georeferencji czy klasyfikacja zobrażeń satelitarnych (Hejmanowska et al., 2020).

Oprogramowanie QGIS zostało zastosowane do stworzenia zestawu danych referencyjnych poprzez wizualną interpretację ortofotomapy oraz mozaik satelitarnych. QGIS

dostarcza zaawansowane narzędzia do digitalizacji, umożliwiające rysowanie i edytowanie obiektów wektorowych oraz pozwala na przeglądanie danych przestrzennych dostępnych w Internecie za pomocą usług sieciowych, takich jak WMS, WMTS czy XYZ Tiles.

4.4.2 SNAP i Sentinel-1 Toolbox

SNAP (ESA, [2023d](#)), czyli Sentinel Application Platform to platforma oprogramowania rozwijana wspólnie przez firmy Brockmann Consult, SkyWatch i C-S na zlecenie Europejskiej Agencji Kosmicznej (ESA), przeznaczona do naukowego wykorzystania misji optycznych i mikrofalowych Sentinel (ESA, [2023e,c](#)). Oprogramowanie SNAP zawiera zestawy narzędzi do wizualizacji, przetwarzania oraz analizy danych teledetekcyjnych, a zaimplementowane narzędzie do przetwarzania grafów (ang. *Graph Processing Tool*, GPT) daje możliwość tworzenia łańcuchów procesów przetwarzania danych zdefiniowanych przez użytkownika (Hejmanowska et al., [2020](#); Moskolaï et al., [2022](#)). Struktura przetwarzania grafów (ang. *Graph Processing Framework*, GPF) w oprogramowaniu SNAP służy do wsadowego przetwarzania danych za pośrednictwem języka Extensible Markup Language (XML) (Moskolaï et al., [2022](#)).

Przetwarzanie danych pochodzących z misji Sentinel-1 umożliwia zestaw narzędzi S1TBX (ESA, [2023a](#)), przeznaczony do przetwarzania danych radarowych. Zestaw narzędzi Sentinel-1 Toolbox (S1TBX) zawiera narzędzia do kalibracji, filtrowania plamek (tzw. efektu pieprzu i soli), koregistracji, ortorektyfikacji, mozaikowania, konwersji danych, polarymetrii i interferometrii (ESA, [2023b](#)). Sentinel-1 Toolbox jest opracowywany dla ESA przez firmę Array we współpracy z DLR, Brockmann Consult i OceanDataLab (ESA, [2023b](#)).

4.4.3 Środowisko języka R

Czynności związane z końcowym przygotowaniem danych wejściowych oraz bezpośrednio z uczeniem maszynowym zostały wykonane z wykorzystaniem środowiska języka R (R Core Team, [2023](#)). R to wieloplatformowy język programowania o otwartym kodzie

źródłowym do obliczeń statystycznych i wizualizacji danych (Lovelace et al., 2019). Dzięki dużej liczbie pakietów R obsługuje również statystyki geoprzestrzenne, modelowanie oraz wizualizację danych przestrzennych (Lovelace et al., 2019). W pracy wykorzystane zostało zintegrowane środowisko programistyczne (ang. *Integrated Development Environment*, IDE) RStudio (RStudio Team, 2020) przeznaczone dla języka R. Poza standardowymi możliwościami środowiska R, w procesie pracy wykorzystane zostały pakiety stworzone przez społeczność R w celu rozszerzenia funkcjonalności tego języka. Do operacji na danych rastrowych zastosowano pakiet *terra* (Hijmans, 2023), natomiast do przetwarzania danych wektorowych używany był pakiet *sf* (Pebesma, 2023). Obliczanie tekstury obrazu Sum Average wyprowadzonej z macierzy współwystępowania poziomu szarości (ang. *gray-level co-occurrence matrix*, GLCM) zostało wykonane przy pomocy pakietu *GLCMTextures* (Ilich, 2023). Losowe generowanie danych przestrzennych umożliwia pakiet *spatstat.random* (Baddeley et al., 2023b) z rodziny pakietów *spatstat* (Baddeley et al., 2023a). Do przeprowadzenia analizy oraz predykcji opartej o elementy uczenia maszynowego wykorzystano pakiet *mlr3* (Lang et al., 2023), w ramach którego użyty został algorytm lasów losowych zaimplementowany w pakiecie *ranger* (Wright et al., 2023). Do obliczeń związanych z teksturami obrazu oraz uczeniem maszynowym wykorzystano pakiet *future* (Bengtsson, 2023), umożliwiający równoległe (wielowątkowe) przetwarzanie wyrażeń R, skracające czas realizacji zadań w stosunku do przetwarzania sekwencyjnego. W funkcji pozwalającej na równoległe obliczanie tekstur obrazu wykorzystano operator `%>%` z pakietu *dplyr* (Wickham et al., 2023), który daje możliwość przekazywania wyniku jednej operacji do następnej.

Rozdział 5

Wyniki

Część **Wyniki** może składać się z jednego lub więcej rozdziałów. Każdy z tych rozdziałów powinien mieć tytuł adekwatny do swojej treści.

Rozdziały wynikowe powinny korzystać z wiedzy opisanej w poprzednich rozdziałach (Rozdziały 2, 3, 4). W przypadku prac analitycznych, ich treść powinna przedstawiać kolejne etapy eksploracji i analizy danych. W przypadku prac technicznych, treść tych rozdziałów powinna opisywać stworzone narzędzia, a następnie pokazywać ich zastosowanie/a.

W przypadku prac technicznych warto pokazywać fragmenty napisanego rozwiązania lub jego wywołania używając bloków kodu.

```
moja_funkcja = function(x){  
  cat(x, "rządzi!")  
}  
moja_funkcja("Autor tej pracy")
```

Autor tej pracy rządzi!

Rozdział 6

Podsumowanie

Podsumowanie pracy jest w pewnym sensie znacznie rozbudowanym abstraktem. Należy wyliczyć i opisać osiągnięcia uzyskane w pracy dyplomowej. Tutaj jednak (w przeciwieństwie do np. rozdziału 1) należy przechodzić od szczegółu do ogółu - co zostało stworzone/określone, jak zostało to zrobione, jakie ma to konsekwencje, itd.

Ten rozdział powinien też zawierać opis kwestii, których nie udało się rozwiązać w pracy dyplomowej (i dlaczego się nie udało) oraz pomysły na przyszłe ulepszenie uzyskanych wyników lub dalsze badania.

Bibliografia

- Baddeley, A, R Turner, and E Rubak (2023a). *spatstat: Spatial Point Pattern Analysis, Model-Fitting, Simulation, Tests*. R package version 3.0-7. <http://spatstat.org/>.
- Baddeley, A, R Turner, E Rubak, and T Davies (2023b). *spatstat.random: Random Generation Functionality for the spatstat Family*. R package version 3.2-1. <http://spatstat.org/>.
- Belgiu, M and L Drăguț (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing* **114**, 24–31.
- Bengtsson, H (2023). *future: Unified Parallel and Distributed Processing in R for Everyone*. R package version 1.33.0. <https://CRAN.R-project.org/package=future>.
- Braun, A (2020). *SAR-based Landcover Classification with Sentinel-1 GRD Products*. Issued October 2020. SkyWatch Space Applications Inc. <http://step.esa.int>.
- Breiman, L (2001). Random Forests. *Machine Learning* **45**, 5–32.
- Drusch, M, U Del Bello, S Carlier, O Colin, V Fernandez, F Gascon, B Hoersch, C Isola, P Laberinti, P Martimort, A Meygret, F Spoto, O Sy, F Marchese, and P Bargellini (2012). Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment* **120**. The Sentinel Missions - New Opportunities for Science, 25–36.
- European Space Agency (2015). *Sentinel-2 User Handbook*. 64 pages. <https://sentinels.copernicus.eu/web/sentinel/user-guides/document-library/-/asset%5C%5Fpublisher/xls%5C%5Fcontent/sentinel-2-user-handbook>.
- European Space Agency (5, 2023a). *ESA Sentinel-1 Toolbox*. Version 9.0.4. <http://step.esa.int/>.
- European Space Agency (5, 2023b). *Sentinel-1 Toolbox*. Accessed on 2023-12-05. <https://sentinel.esa.int/web/sentinel/toolboxes/sentinel-1>.

- European Space Agency (2023c). *SNAP*. Accessed on 2023-12-05. <https://step.esa.int/main/toolboxes/snap/>.
- European Space Agency (5, 2023d). *SNAP - ESA Sentinel Application Platform*. Version 9.0.8. <http://step.esa.int>.
- European Space Agency (5, 2023e). *SNAP Desktop*. Accessed on 2023-12-05. <https://github.com/senbox-org/snap-desktop>.
- European Space Agency (n.d.[a]). *Sentinel-1 Acquisition Modes - Stripmap*. Accessed on 2023-12-08. <https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-1-sar/acquisition-modes/stripmap>.
- European Space Agency (n.d.[b]). *Sentinel-1 Data Products*. Accessed on 2023-12-08. <https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-1/data-products>.
- European Space Agency (n.d.[c]). *Sentinel-1 Instrument Payload*. Accessed on 2023-12-08. <https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-1/instrument-payload>.
- European Space Agency (n.d.[d]). *Sentinel-1 Mission Objectives*. Accessed on 2023-12-08. <https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-1/mission-objectives>.
- European Space Agency (n.d.[e]). *Sentinel-1 Mission Overview*. Accessed on 2023-12-08. <https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-1/overview>.
- Filipponi, F (2019). Sentinel-1 GRD Preprocessing Workflow. *Proceedings* **18**(1).
- Flenniken, JM, S Stuglik, and BV Iannone (2020). Quantum GIS (QGIS): An introduction to a free alternative to more costly GIS platforms. *EDIS* **2020**(2), 7.
- Gislason, PO, JA Benediktsson, and JR Sveinsson (2006). Random Forests for land cover classification. *Pattern Recognition Letters* **27**(4). *Pattern Recognition in Remote Sensing (PRRS 2004)*, 294–300.
- Główny Urząd Geodezji i Kartografii (n.d.). *Ortofotomapa (ORTO)*. Accessed on 2023-12-09. <https://www.geoportal.gov.pl/pl/dane/ortofotomapa-orto/>.

- Gong, P, DJ Marceau, and PJ Howarth (1992). A comparison of spatial feature extraction algorithms for land-use classification with SPOT HRV data. *Remote Sensing of Environment* **40**(2), 137–151.
- GRASS Development Team (2023). *r.texture - GRASS GIS manual*. Accessed on 2023-12-11. <https://grass.osgeo.org/grass83/manuals/r.texture.html>.
- Hall-Beyer, M (2017). *GLCM Texture: A Tutorial v. 3.0 March 2017*. <https://doi.org/10.13140/RG.2.2.12424.21767>.
- Haralick, RM, K Shanmugam, and IH Dinstein (1973). Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics* (6), 610–621.
- Hejmanowska, B and P Wezyk (2020). *Dane satelitarne dla administracji publicznej*. Polska Agencja Kosmiczna.
- Hijmans, RJ (2023). *terra: Spatial Data Analysis*. R package version 1.7-55. <https://rspatial.org/>.
- Ilich, A (2023). *GLCMTextures: GLCM Textures of Raster Layers*. R package version 0.4, <https://github.com/ailich/GLCMTextures>. <https://ailich.github.io/GLCMTextures/>.
- Ismail, MH, H Pakhriazad, and M Shahrin (2009). Evaluating supervised and unsupervised techniques for land cover mapping using remote sensing data. *Geografia : Malaysian Journal of Society and Space*.
- Kupidura, P (2019). The Comparison of Different Methods of Texture Analysis for Their Efficacy for Land Use Classification in Satellite Imagery. *Remote Sensing* **11**(10).
- Lang, M, B Bischl, J Richter, P Schratz, M Binder, F Pfisterer, R Sonabend, and M Becker (2023). *mlr3: Machine Learning in R - Next Generation*. R package version 0.17.0. <https://CRAN.R-project.org/package=mlr3>.
- Lesiv, M, L See, JC Laso Bayas, T Sturn, D Schepaschenko, M Karner, I Moorthy, I McCallum, and S Fritz (2018). Characterizing the Spatial and Temporal Availability of Very High Resolution Satellite Imagery in Google Earth and Microsoft Bing Maps as a Source of Reference Data. *Land* **7**(4).
- Lewiński S.; Aleksandrowicz, S (2012). Ocena możliwości wykorzystania tekstury w rozpoznaniu podstawowych klas pokrycia terenu na zdjęciach satelitarnych różnej rozdzielczości. *Archiwum Fotogrametrii, Kartografii i Teledetekcji* **23**, 229–237.

- Lovelace, R, J Nowosad, and J Muenchow (2019). *Geocomputation with R*. CRC Press.
- Mahesh, B (2019). Machine Learning Algorithms - A Review. en. 9(1).
- Main-Knorn, M, B Pflug, J Louis, V Debaecker, U Müller-Wilm, and F Gascon (2017). Sen2Cor for Sentinel-2. In: *Image and Signal Processing for Remote Sensing XXIII*. Ed. by L Bruzzone. Vol. 10427. International Society for Optics and Photonics. SPIE, pp.37–48. <https://doi.org/10.1117/12.2278218>.
- Malinowski, R, S Lewiński, M Rybicki, E Gromny, M Jenerowicz, M Krupiński, A Nowakowski, C Wojtkowski, M Krupiński, E Krätzschmar, and P Schauer (2020). Automated Production of a Land Cover/Use Map of Europe Based on Sentinel-2 Imagery. *Remote Sensing* 12(21).
- Maxwell, AE, TA Warner, and F Fang (2018). Implementation of machine-learning classification in remote sensing: an applied review. *International Journal of Remote Sensing* 39(9), 2784–2817. eprint: <https://doi.org/10.1080/01431161.2018.1433343>.
- Microsoft (2020). *Maps and Geolocation Data*. Bing Maps API, SDKs and Services. Accessed on 2023-12-09. <https://www.microsoft.com/en-us/maps/bing-maps/location-data>.
- Moskolai, WR, W Abdou, A Dipanda, and Kolyang (2022). A Workflow for Collecting and Preprocessing Sentinel-1 Images for Time Series Prediction Suitable for Deep Learning Algorithms. *Geomatics* 2(4), 435–456.
- Mumby, PJ and AJ Edwards (2002). Mapping marine environments with IKONOS imagery: enhanced spatial resolution can deliver greater thematic accuracy. *Remote Sensing of Environment* 82(2), 248–257.
- Pebesma, E (2023). *sf: Simple Features for R*. R package version 1.0-13. <https://CRAN.R-project.org/package=sf>.
- Phiri, D, M Simwanda, S Salekin, VR Nyirenda, Y Murayama, and M Ranagalage (2020). Sentinel-2 Data for Land Cover/Use Mapping: A Review. *Remote Sensing* 12(14).
- Plakman, V, J Rosier, and J van Vliet (2022). Solar park detection from publicly available satellite imagery. *GIScience & Remote Sensing* 59(1), 462–481. eprint: <https://doi.org/10.1080/15481603.2022.2036056>.

- Planet Labs Inc. (2019). *Planet Basemaps Product Specification*. Accessed on 2023-12-05. © Planet Labs Inc. 2019. All rights reserved. <https://assets.planet.com/products/basemap/planet-basemaps-product-specifications.pdf>.
- QGIS Development Team (2023). *QGIS Geographic Information System*. QGIS Association. <https://www.qgis.org>.
- R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. <https://www.R-project.org/>.
- RStudio Team (2020). *RStudio: Integrated Development Environment for R*. RStudio, PBC. Boston, MA. <http://www.rstudio.com/>.
- Sarker, I (2021). *Machine Learning: Algorithms, Real-World Applications and Research Directions*.
- Sekulić, A, M Kilibarda, GB Heuvelink, M Nikolić, and B Bajat (2020). Random Forest Spatial Interpolation. *Remote Sensing* **12**(10).
- Sheykhoumou, RM and M Mahdianpari (2020). Support Vector Machine vs. Random Forest for Remote Sensing Image Classification: A Meta-analysis and systematic review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
- Sindayigaya, L and A Dey (2022). Machine Learning Algorithms: A Review. *Information Systems Journal*, vol. ISJ-RA-3392, sp 6.
- Tucker, CJ (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment* **8**(2), 127–150.
- Wang, J, J Liu, and L Li (2022). Detecting Photovoltaic Installations in Diverse Landscapes Using Open Multi-Source Remote Sensing Data. *Remote Sensing* **14**(24).
- Wickham, H, R François, L Henry, K Müller, and D Vaughan (2023). *dplyr: A Grammar of Data Manipulation*. R package version 1.1.4. <https://CRAN.R-project.org/package=dplyr>.
- Wright, MN, S Wager, and P Probst (2023). *ranger: A Fast Implementation of Random Forests*. R package version 0.16.0. <https://CRAN.R-project.org/package=ranger>.
- Xu, H (2006). Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing* **27**(14), 3025–3033. eprint: <https://doi.org/10.1080/01431160600589179>.

- Zha, Y, J Gao, and S Ni (2003). Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *International Journal of Remote Sensing* **24**(3), 583–594. eprint: <https://doi.org/10.1080/01431160304987>.
- Zhang, X, M Zeraatpisheh, MM Rahman, S Wang, and M Xu (2021). Texture Is Important in Improving the Accuracy of Mapping Photovoltaic Power Plants: A Case Study of Ningxia Autonomous Region, China. *Remote Sensing* **13**(19).