



Kierunek studiów: Geoinformacja

Nr albumu: 461791

Filip Ratajszczak

**Wykrywanie farm fotowoltaicznych na podstawie
danych teledetekcyjnych**

*Detection of photovoltaic farms based on remote
sensing data*

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

Streszczenie

Abstrakt

Streszczenie powinno przedstawać skrótnie główny problem pracy i jego rozwiązanie. Możliwa struktura streszczenia to: (1) 1-3 zdania wstępne 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
3 Materiały	9
3.1 Zdjęcia satelitarne	10
3.2 Ortofotomapa i mozaiki zdjęć satelitarnych	14
4 Metody	17
4.1 Przygotowanie danych	17
4.2 Próbki treningowe i testowe	23
4.3 Przetwarzanie końcowe	24
4.4 Uczenie maszynowe	26
4.5 Ważność zmiennych	32
4.6 Oprogramowanie	33
5 Wyniki	37
5.1 Ocena jakości modeli	37
5.2 Ważność zmiennych	37
5.3 Wizualna kontrola wyników klasyfikacji	37
6 Podsumowanie	39
Bibliografia	41

Rozdział 1

Wprowadzenie

Wprowadzenie powinno mieć charakter opisu od ogólnego do szczegółów (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

Rozdział 3

Materiały



Rycina 3.1: Obszar badań

3.1 Zdjęcia satelitarne

3.1.1 Sentinel-1

Sentinel-1 to wspólna inicjatywa Komisji Europejskiej i Europejskiej Agencji Kosmicznej w ramach programu Copernicus, mająca na celu 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, 2023g). 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, 2023g).

Sentinel-1 posiada pojedynczy radar z syntetyzowaną aperturą (ang. *Synthetic Aperture Radar*, SAR), który działa na jednej częstotliwości środkowej 5,405 GHz, co odpowiada długości fali 5,6 cm (Braun, 2020; ESA, 2023e). Częstotliwość radaru Sentinel-1, mieszcząca się w pasmie C (od 4 do 8 GHz), determinuje jego zdolność do penetracji jedynie górnych warstw koron drzew, roślinności oraz gleby (ESA, 2023j). Radar jest najczęściej wykorzystywany 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 syntetyzowaną aperturą (SAR) umożliwia prowadzenie obserwacji zarówno w nocy, jak i w dzień, niezależnie od warunków pogodowych, przez co czas rewizyty konstelacji dwóch satelitów dla obszaru Polski wynosi około 2 dni (Attema et al., 2008; ESA, 2023h). W grudniu 2021 roku w satelicie Sentinel-1B wystąpiła anomalia związana z zasilaniem elektroniki radaru, uniemożliwiająca dalsze dostarczanie danych radarowych (ESA, 2023b). W wyniku awarii satelita Sentinel-1B został wyłączony z użytku, a wystrzelenie platformy Sentinel-1C zaplanowano na marzec 2024 roku (ESA, 2023b; eoPortal, 2023).

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). Interferometria radarowa (ang. *Interferometric Synthetic Aperture Radar*, InSAR) to technika umożliwiająca generowanie cyfrowych modeli wysokościowych (ang. *digital elevation model*, DEM) oraz pomiary deformacji terenu na podstawie informacji o fazie sygnału (Hanssen, 2001; Hejmanowska et al., 2020). SLC to obrazy SAR w geometrii ukośnej, posiadające informację o wartości amplitudy i fazie 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

Tabela 3.1: Tryby radaru Sentinel-1 (ESA, 2023)

Tryb	Kąt padania wiązki [°]	Rozdzielcość przestrzenna [m]	Szerokość rejestrowanego pasa [km]	Polaryzacje (H=pozioma, V=pionowa)
Stripmap	20 - 45	5 x 5	80	HH+HV, VH+VV, HH, VV
Interferometric Wide Swath	29 - 46	5 x 20	250	HH+HV, VH+VV, HH, VV
Extra Wide Swath	19 - 47	20 x 40	400	HH+HV, VH+VV, HH, VV
Wave	22 - 35 35 - 38	5 x 5	20 x 20	HH, VV

tracimy informację fazową sygnału, co sprawia, że produkty GRD nie są odpowiednie do interferometrii radarowej (ESA, 2023d).

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 oraz Wave (WV) - tryb obrazowania oceanów (Hejmanowska et al., 2020; ESA, 2023e,c). Charakterystykę konkretnych trybów akwizycji radaru Sentinel-1 przedstawia tabela 3.1 (ESA, 2023f).

W pracy wykorzystano dwie polaryzacje (VV i VH) pochodzące z produktów Ground Range Detected (GRD), zarejestrowanych w trybie Interferometric Wide Swath (IW) przez platformę Sentinel-1A w dniu 8 maja 2023 roku. Użyty zestaw danych został utworzony poprzez połączenie dwóch sąsiednich produktów Sentinel-1 GRD, a następnie ograniczenie obszaru analizy do kafla Sentinel-2 o oznaczeniu 33UWV.

3.1.2 Senitnel-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. Czas rewizyty konstelacji dwóch satelitów wzrasta z 5 dni nad równikiem do 2-3 dni na średnich szerokościach geograficznych, obejmujących obszar Polski (Hejmanowska et al., 2020; ESA, 2023k).

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 km^2 ($100 \times 100 \text{ km}$) 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 3.2 (ESA, 2023l)) wykorzystano 10 zakresów, ponieważ pasma rejestrowane w rozdzielczości 60 metrów są przeznaczone głównie do korekcji atmosferycznych i detekcji chmur. Kanał 1 (443 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).

Tabela 3.2: Kanały spektralne satelitów Sentinel-2 (ESA, 2023)

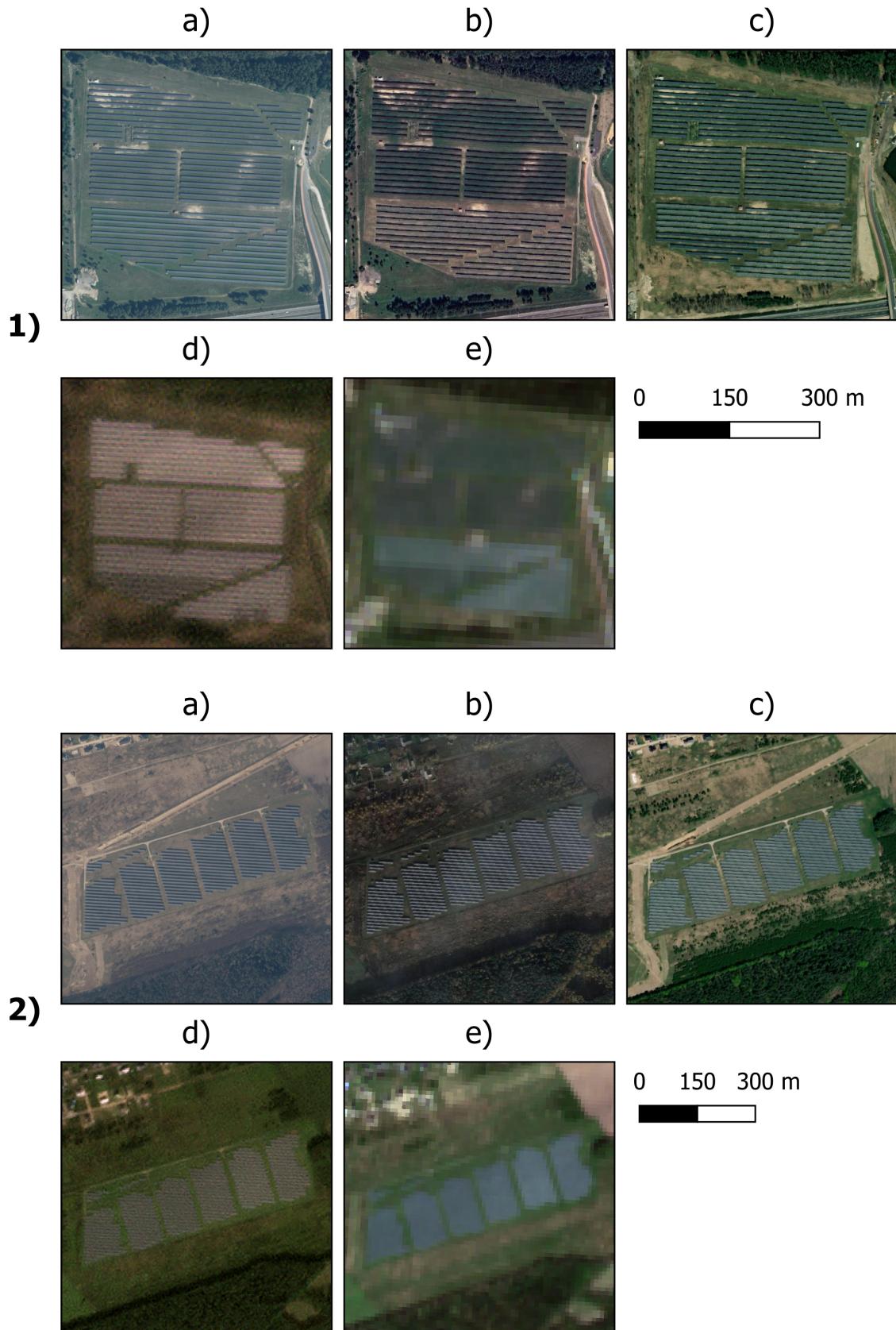
Kanał	Nazwa kanału	Centralna długość fali [nm]	Zakres spektralny [nm]	Rozdzielcość przestrzenna [m]
B01	Coastal Aerosol	443	433–453	60
B02	Blue	493	458–523	10
B03	Green	560	543–578	10
B04	Red	665	650–680	10
B05	Vegetation RedEdge	704	698–713	20
B06	Vegetation RedEdge	740	733–748	20
B07	Vegetation RedEdge	783	773–793	20
B08	NIR	833	785–900	10
B8A	NIR	865	855–875	20
B09	Water Vapour	945	935–955	60
B10	Cirrus	1374	1360–1390	60
B11	SWIR	1610	1565–1655	20
B12	SWIR	2190	2100–2280	20

3.2 Ortofotomapa i mozaiki zdjęć satelitarnych

Do lokalizacji oraz wektoryzacji istniejących farm fotowoltaicznych wykorzystano ortofotomapę udostępnianą przez Główny Urząd Geodezji i Kartografii (GUGiK) 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 stworzenie zbioru danych testowych i treningowych wykorzystano ortomozaiki Google Satellite, Bing Aerial oraz Planet Basemaps, udostępniane w formie usług sieciowych (WMS, WMTS, XYZ Tiles). Mozaiki te są tworzone na podstawie komercyjnych zdjęć satelitarnych wykonywanych przez podmioty takie jak Maxar Technologies, Airbus czy Planet Labs.

Większość ortofotomap dostarczanych przez GUGiK jest zrealizowana w standardzie 25 x 25 cm, jednakże na obszarach miejskich charakteryzuje się one rozdzielcością przestrzenną wynoszącą 10 cm lub nawet wyższą (GUGiK, 2023). Rozdzielcość przestrzenna wykorzystanych mozaik obrazów satelitarnych (oprócz Planet Basemaps) jest wyższa niż 1 m, na przykład mozaika Bing Aerial dostarczana przez firmę Microsoft cechuje się wielkością komórki od 30 do 60 cm (Microsoft, 2023). Często jednak nie jest

możliwie 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, co umożliwia generowanie wysokorozdzielczych mozaik, które są dokładne radiometrycznie i przestrzennie, a także charakteryzują się zminimalizowanym wpływem czynników atmosferycznych (Planet Labs Inc., 2019).



Rycina 3.2: Porównanie wyglądu farm fotowoltaicznych w Kulicach k. Nowogardu (1) i Wałczu (2) na ortofotomapie udostępnianej przez GUGiK (a), Google Satellite (b), Bing Aerial (c), Planet Basemaps (d) oraz w kompozycji RGB Sentinel-2 (e)

Rozdział 4

Metody

WSTĘP

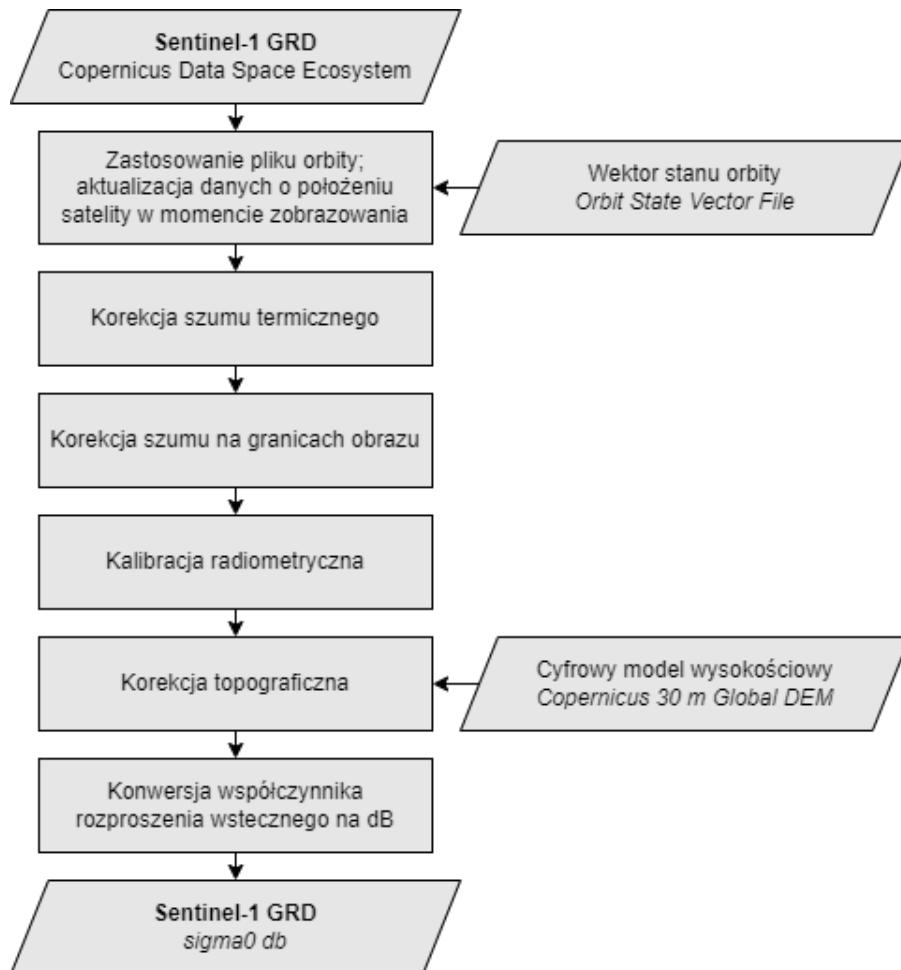
RYCINA - OGÓLNY SCHEMAT PRZEBIEGU BADANIA (WORKFLOW)

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 Filippone (2019), obejmujący:

1. aktualizację informacji o położeniu satelity w momencie zobrazowania poprzez pobranie dokładnych wektorów stanu orbity dla produktu zapewniając precyzyjne informacje o pozycji i prędkości satelity podczas akwizycji;
2. korekcję szumów termicznych;
3. korekcję szumów na granicach obrazów;
4. obliczenie współczynnika rozproszenia wstecznego (ang. *backscatter coefficient*) sigma0 za pomocą kalibracji radiometrycznej;



Rycina 4.1: Przebieg wstępniego przetwarzania danych *Sentinel-1 Ground Range Detected (GRD)*

5. korekcję topograficzną (ortorektyfikacja za pomocą *Copernicus 30 m Global DEM*);
6. konwersję współczynnika rozproszenia wstecznego na dB za pomocą transformacji logarytmicznej

Dane *Sentinel-1 GRD* dla obu polaryzacji (VV i VH) przygotowano w sposób przedstawiony na rycinie 4.1 przy użyciu narzędzi dostępnych w *ESA Sentinel-1 Toolbox (S1TBX)* ([ESA, 2023a](#)), a proces ten został ułożony w sekwencję operacji przetwarzania za pomocą narzędzia *GraphBuilder* w oprogramowaniu *SNAP* ([ESA, 2023n](#)). 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



Rycina 4.2: Obszar badań i zasięg wykorzystanych produktów Sentinel-1

odpowiednio ograniczone do obszaru zainteresowania. Na granicy sąsiednich produktów Sentinel-1 GRD występowała przestrzeń bez danych o szerokości jednej komórki, co wymagało wypełnienia tego obszaru danymi przy użyciu funkcji `focal` z pakietu `terra` (Hijmans, 2023).

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. Przepróbkowanie (ang. *resampling*) zostało przeprowadzone przy pomocy funkcji `resample` z pakietu `terra` (Hijmans, 2023), stosując interpolację dwuliniową (ang. *bilinear interpolation*).

W teledetekcji miarą opisującą wielkość odbicia jest rzeczywisty współczynnik odbicia, nazywany również reflektancją, którego wartości mieszczą się w przedziale od 0 do 1 (Hejmanowska et al., 2020). Informacje dotyczące odbitego promieniowania w przypadku

danych Sentinel-2 są udostępniane za pomocą współczynnika odbicia, wyrażonego w zakresie od 0 do 10000, gdzie wartości te reprezentują jasność komórki, określana w języku angielskim jako *Digital Number* (DN). Po standaryzacji wszystkich kanałów Sentinel-2 do wspólnej rozdzielczości, dokonano przeliczenia wartości kanałów w celu uzyskania rzeczywistego współczynnika odbicia, poprzez podzielenie wartości DN przez 10000.

4.1.3 Wskaźniki spektralne

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

- znormalizowany różnicowy wskaźnik wegetacji (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* – reflektancja w kanale bliskiej podczerwieni, *Red* – reflektancja w kanale czerwonym

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

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

, gdzie *SWIR1* – reflektancja w kanale średniej podczerwieni, *NIR* – reflektancja w kanale bliskiej podczerwieni

- znormalizowany zmodyfikowany różnicowy wskaźnik 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* – reflektancja w kanale zielonym, *SWIR1* – reflektancja w kanale średniej podczerwieni

4.1.4 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 Development Team, 2023). W zależności od zastosowanej funkcji wybrane cechy obrazu zostają uwidoczone w porównaniu z jego obrazem wejściowym (Lewiński, 2012). Informacja o tekście 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 Development Team, 2023). Można ją opisać za pomocą tonu (intensywność poziomu szarości) i struktury (relacje przestrzenne) (GRASS Development Team, 2023). Model 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 odstępy między komórkami (Kupidura, 2019).

Przydatność i wykorzystanie tekstury w dużym stopniu zależy od rozdzielczości zdjęć satelitarnych i wielkości zjawiska, które ukształtowało teksturę (GRASS Development Team, 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 wykorzystano jedynie teksturę średniej sumy (ang. *Sum Average*, SA lub SAVG), wskazaną 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. Przeprowadzono obliczenia sześciu tekstur obrazów dla dwóch kanałów Sentinel-2 (B02 i B8A), dwóch wskaźników teledetekcyjnych (NDBI i nMDWI) oraz dla obu polaryzacji Sentinel-1 (VV i VH), zgodnie z wskazaniami Wang et al. (2022). W celu skrócenia czasu obliczeń, w pracy zdecydowano się na zastosowanie ruchomego okna o sąsiedztwie 9 komórek.

4.1.5 Łączenie danych

W celu uzyskania spójnych wielokanałowych rastrów, wszystkie zbiory danych zostały sprowadzone do wspólnej rozdzielczości i siatki. Rozdzielcość przestrzenna danych Sentinel-1 GRD, analogicznie do danych Sentinel-2 przygotowanych w sposób przedstawiony w sekcji 4.1.2 wynosi 10 m. Mimo że rozdzielcość przestrzenna danych Sentinel-1 GRD jest identyczna z rozdzielcością danych Sentinel-2, siatki przestrzenne obu zbiorów różniły się od siebie, przez co wymagana była transformacja danych Sentinel-1 do siatki danych Sentinel-2 w celu zachowania ich zgodności. Przetransformowane dane zostały wykorzystane do obliczeń tekstur obrazu oraz wskaźników teledetekcyjnych. Po uzyskaniu produktów pochodnych, w zależności od wariantu, dane zostały scalone w formie kilku wielokanałowych rastrów, które posłużyły do wyodrębnienia zestawu danych

Tabela 4.1: Warianty [NAZWA DO ZMIANY!!!]

Wariant	Liczba zmiennych	Zmienne ^a
1	10	S2: B03, B04, B05, B06, B07, B08, B8A, B09, B10, B11, B12
2	13	S2: B03, B04, B05, B06, B07, B08, B8A, B09, B10, B11, B12; wskaźniki spektrelne: NDVI, NDBI, mNDWI
3	16	S2: B03, B04, B05, B06, B07, B08, B8A, B09, B10, B11, B12; wskaźniki spektrelne: NDVI, NDBI, mNDWI; tekstury: B02_SAVG, B8A_SAVG, NDBI_SAVG, mNDWI_SAVG
4	12	S2: B03, B04, B05, B06, B07, B08, B8A, B09, B10, B11, B12; S1: VV, VH
5	16	S2: B03, B04, B05, B06, B07, B08, B8A, B09, B10, B11, B12; S1: VV, VH; tekstury: 02_SAVG, B8A_SAVG, VV_SAVG, VH_SAVG
6	21	S2: B03, B04, B05, B06, B07, B08, B8A, B09, B10, B11, B12; wskaźniki spektrelne: NDVI, NDBI, mNDWI; S1: VV, VH; tekstury: B02_SAVG, B8A_SAVG, VV_SAVG, VH_SAVG, NDBI_SAVG, mNDWI_SAVG

^a Uwaga: S2 oznacza Sentinel-2, podczas gdy S1 oznacza Sentinel-1

treningowych oraz do przeprowadzenia predykcji. Warianty zestawów danych zostały szczegółowo przedstawione w tabeli 4.1.

4.2 Próbki treningowe i testowe

Na podstawie ortofotomapy oraz mozaik satelitarnych wskazanych w sekcji 3.2 zwektryzowane zostały prawdopodobnie wszystkie farmy fotowoltaiczne na obszarze kafla Sentinel-2 o oznaczeniu 33UWV istniejące w czasie wykonywania wykorzystanych zobrażeń (8 maja 2023 roku). Z każdego zwektryzowanego poligonu, reprezentującego obszar pod panelami fotowoltaicznymi pozyskano dwie losowo zlokalizowane próbki, stanowiące obserwacje pozytywne. Lokalizacje próbek negatywnych, znajdujących się poza obszarami oznaczonymi jako farmy fotowoltaiczne na obszarze kafla Sentinel-2 o oznaczeniu 33UWV, zostały wylosowane wykorzystując próbkowanie losowe stratyfikowane (ang. *stratified*). Próbkowanie losowe stratyfikowane polega na podziale obszaru analizy na regularne komórki, a następnie losowaniu lokalizacji punktu w każdej komórce (Nowosad, 2021).

Po wykonaniu pierwszych testowych predykcji zauważono, że modele przeuczały się na niektórych typach pokrycia terenu i użytkowania ziemi, wskazując farmy fotowoltaiczne w miejscach, gdzie faktycznie nie występowały. W celu poprawy wyników predykcji dodatkowe lokalizacje negatywnych próbek zostały wylosowane na obszarach niepoprawnie sklasyfikowanych, wykorzystując dane z OpenStreetMap (OSM contributors, 2024). OpenStreetMap (OSM) to bezpłatna, otwarta geograficzna baza danych aktualizowana i utrzymywana przez społeczność wolontariuszy (Bennett, 2010).

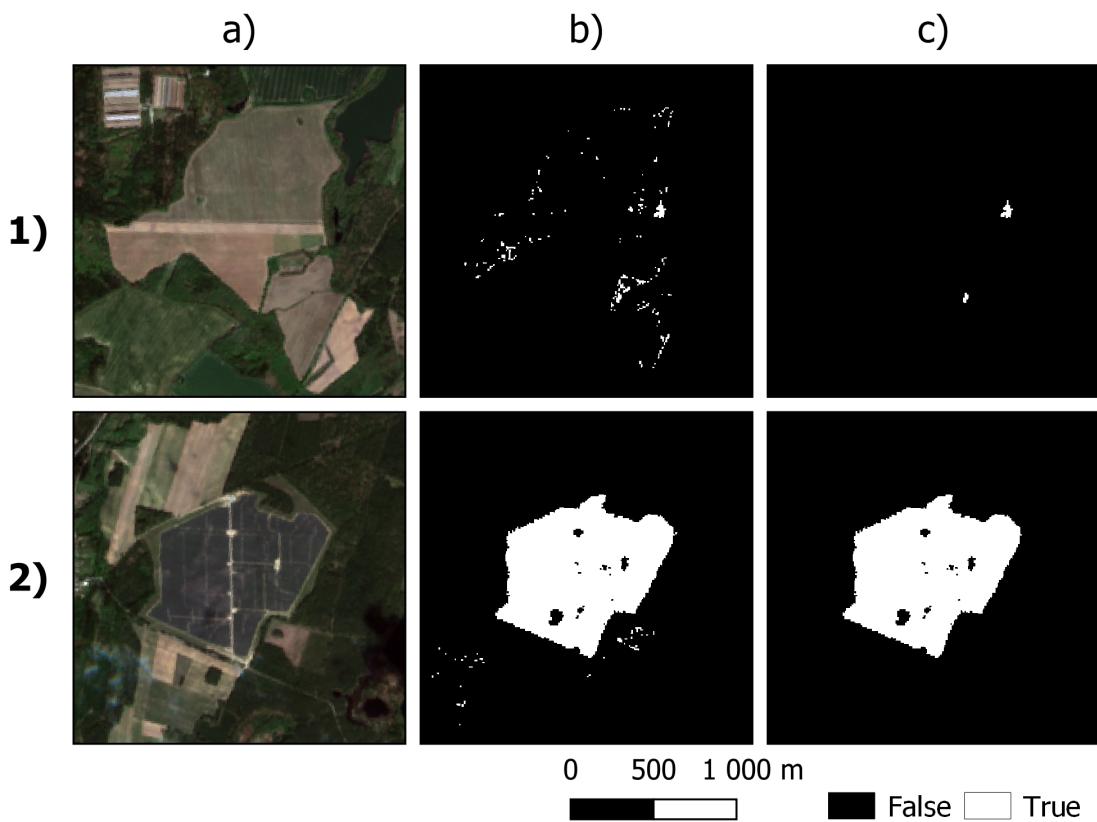
Obszary niepoprawnie sklasyfikowane przez pierwsze modele to plaże, budynki oraz drogi. W celu poprawy wyników predykcji pobrano z bazy danych OpenStreetMap dane przestrzenne o plażach (tag:natural=beach), budynkach (key:building) oraz drogach (key:highway) na obszarze kafla Sentinel-2 o oznaczeniu 33UWV, z których wylosowano lokalizacje kolejnych negatywnych próbek. Dodatkowo, lokalizacje negatywnych obserwacji zostały wylosowane na zbiornikach wodnych (jeziorach; tag:water=lake).

Przy tworzeniu kolejnych predykcji zauważono również skłonność do regularnego przeuczania się kolejnych, poprawionych modeli na terenach oznaczonych w OSM jako kopalnie torfu. W celu eliminacji tych błędnie sklasyfikowanych terenów, przy tworzeniu ostatecznych modeli, wykorzystano również negatywne próbki wylosowane na terenach kopalni odkrywkowych oznaczonych w bazie OpenStreetMap jako tag:landuse=quarry.

Dla każdej wylosowanej próbki zostały wyekstraktowane wartości pochodzące z danych teledetekcyjnych i ich pochodnych, przygotowane zgodnie z opisem przedstawionym w sekcjach 4.1.1, 4.1.2, 4.1.3 oraz 4.1.4. Tak przygotowane próbki były następnie wykorzystane przy tworzeniu modeli uczenia maszynowego umożliwiających wykrywanie farm fotowoltaicznych na podstawie danych teledetekcyjnych i ich pochodnych.

4.3 Przetwarzanie końcowe

Ostatnim etapem procesu detekcji farm fotowoltaicznych było określenie metod przetwarzania końcowego (ang. *post-processing*), mających na celu udoskonalenie wyników wykrywania farm fotowoltaicznych. Z teoretycznego punktu widzenia, obszar zajmowany przez farmę fotowoltaiczną lub jej fragmenty powinien być większy niż powierzchnia



Rycina 4.3: POSTPROCESSING [DO ZMIANY]

pojedynczej komórki lub grupy kilku komórek. W celu usunięcia pozytywnych predykacji, które reprezentowały pojedyncze komórki lub obszary składające się z 10 lub mniej komórek, o powierzchni mniejszej/równiej 1000 m^2 , zastosowano sekwencję procesów przetwarzania, dzięki której udało się pozbyć tzw. efektu soli i pieprzu (ang. *salt-and-pepper effect*).

Przetwarzanie końcowe obejmowało wektoryzację rastrowych wyników predykci dla każdego wariantu, a następnie selekcję pozytywnych predykacji oraz obliczenie powierzchni każdego wybranego obszaru. Na podstawie obliczonej powierzchni usunięto wszystkie pozytywne predykce reprezentujące pojedyncze komórki lub obszary składające się z 10 lub mniej komórek, dzięki czemu wiele błędnych predykacji zostało usuniętych. W rezultacie uzyskano końcowy produkt wykrywania farm fotowoltaicznych oparty na danych teledetekcyjnych w formie wektorowej.

W celu stworzenia wizualizacji oraz porównania produktów końcowych z predykcjami niepoddanymi etapowi przetwarzania końcowego, produkty wektorowe zostały ponownie przetworzone do formy rastrowej.

4.4 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 wykorzystywanyimi 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 przejrzeniu danych (Mahesh, 2019). Uczenie maszynowe jest wykorzystywane w celu doskonalenia efektywnego przetwarzania danych przez maszyny (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 stosowanie kilku różnych algorytmów uczenia maszynowego do klasyfikacji obrazów satelitarnych (Sheykhou et al., 2020), które zazwyczaj wykorzystują 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 (redukci cech) oraz identyfikacji skojarzeń i relacji (Sarker, 2021).

Nadzorowane algorytmy uczenia maszynowego wykorzystują oznaczone dane treningowe do znajdywania powiązań pomiędzy różnymi zmiennymi. Proces uczenia nadzorowanego zachodzi, gdy określone cele mają zostać osiągnięte na podstawie konkretnego zestawu danych wejściowych (treningowych). W uczeniu nadzorowanym wyróżniamy klasyfikację, która dzieli dane na klasy, oraz regresję, która umożliwia oszacowanie określonych wartości (Sarker, 2021).

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

4.4.1 Metoda lasów losowych

Random Forest stał się jednym z najpopularniejszych klasyfikatorów uczenia maszynowego wykorzystywanych w dziedzinie teledetekcji ze względu na wysoką dokładność klasyfikacji i efektywność obliczeniową (Belgiu et al., 2016; Sheykhou et al., 2020). Metoda lasów losowych charakteryzuje się pewną 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 uczenia zespołowego (ang. *ensemble learning*), czyli łączenia wielu słabych 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 dokładne, 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 (Breiman, 2001).

4.4.2 Dostrajanie modeli

Celem optymalizacji hiperparametrów lub dostrajania modelu jest znalezienie optymalnej konfiguracji hiperparametrów algorytmu uczenia maszynowego dla danego zadania (Bischl et al., 2024). Optymalizacja hiperparametrów (ang. *hyperparameters*) odgrywa kluczową rolę w osiągnięciu najwyższej mocy predykcyjnej i jakości modelu (Schratz et al., 2019). Hiperparametry są ustawiane przed rozpoczęciem procesu uczenia, a ich optymalna konfiguracja jest zwykle znajdowana w określonej przestrzeni poszukiwań (ang. *search space*) i ustalana na podstawie zagnieżdzonej walidacji krzyżowej (ang. *nested cross-validation*) (Lovelace et al., 2019). Metoda zagnieżdzonej walidacji krzyżowej to rozwinięcie walidacji krzyżowej, opisanej w sekcji 4.4.3, przeznaczone do procesu strojenia hiperparametrów. Każdy fold utworzony w przestrzennej walidacji krzyżowej jest dzielony na kolejne, zewnętrzne i wewnętrzne foldy. W zewnętrznym obiegu model jest trenowany na danych z jednego obszaru przestrzennego i testowany na innym, podczas gdy wewnętrzny obieg służy do doboru optymalnych hiperparametrów. Cały proces jest następnie powtarzany na każdym z k zewnętrznych foldów, co prowadzi do określenia najlepszego ustawienia hiperparametrów. Wykorzystanie tych samych danych do oceny jakości i dostrajania mogłoby potencjalnie prowadzić do nadmiernie wysokich wyników oceny jakości, czego można uniknąć, stosując zagnieżdzoną przestrzenną kroswalidację. Cały ten proces jest nazywany strojeniem hiperparametrów.

Lasy losowe często wykazują satysfakcjonujące wyniki nawet z domyślnymi wartościami hiperparametrów, co może być jednym z powodów ich dużej popularności (Lovelace et al., 2019). Chociaż dostrojenie lasów losowych powinno poprawiać jakość modeli, korzyści ze strojenia są znacznie mniejsze w porównaniu do innych algorytmów uczenia maszynowego, takich jak maszyny wektorów nośnych (ang. *Support Vector Machines*, SVM) (Probst et al., 2019).

Hiperparametry `mtry`, `sample.fraction` i `min.node.size` są parametrami określającymi stopień losowości lasu losowego i powinny zostać odpowiednio dostrojone (Probst

Tabela 4.2: Optymalne hiperparametry otrzymane w wyniku dostrajania modeli RF

Wariant ^a	Optymalizowane parametry						AUC ^b
	mtry	sample.fraction	min.node.size	num.trees	max.depth		
1	9	0.6843	2	180	36	0.9886	
2	7	0.7373	1	372	23	0.9914	
3	7	0.8847	5	388	75	0.9904	
4	6	0.8758	5	291	54	0.9881	
5	5	0.6898	1	234	99	0.9850	
6	8	0.8847	5	388	75	0.9905	

^a Patrz: tabela 4.1^b Patrz: sekcja 4.3.3

et al., 2019). Liczba losowo wybranych zmiennych `mtry` wskazuje, ile zmiennych predykcyjnych powinno zostać użytych w każdym drzewie (Lovelace et al., 2019). Parametr `sample.fraction` odnosi się do wielkości próbki, czyli ułamka obserwacji użytego w każdym drzewie (Lovelace et al., 2019). Mniejsze frakcje prowadzą do większej różnorodności drzew, a tym samym do mniejszej korelacji między nimi, co pozytywnie wpływa na dokładność predykcji przy agregacji drzew (Probst et al., 2019). Minimalna wielkość węzła `min.node.size` określa minimalną liczbę obserwacji w węźle końcowym (Probst et al., 2019). W ramach optymalizacji uwzględniono również parametry `num.trees` oraz `max.depth`, odnoszące się odpowiednio do liczby drzew w lesie oraz maksymalnej głębokości pojedynczego drzewa.

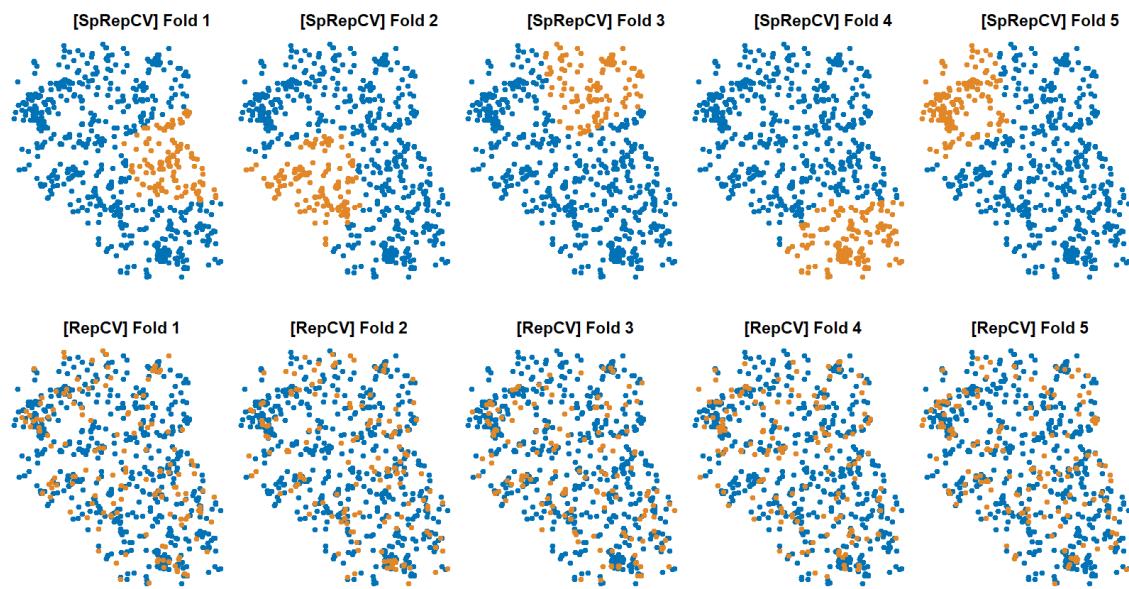
Kombinacje hiperparametrów zostały wybrane losowo, jednak pozostawały w określonych granicach strojenia ustalonych za pomocą pakietu *paradox* (Lang et al., 2023b). Zasięg przestrzeni strojenia został wybrany zgodnie z wartościami zalecanymi w dedykowanym do tego pakietu *mlr3tuningspaces* (Becker, 2023) lub literaturze (Probst et al., 2019; Schratz et al., 2019). `mtry` powinno przyjmować wartości z przedziału od 1 do liczby predyktorów, `sample.fraction` powinno mieścić się w zakresie od 0,2 do 0,9, a `min.node.size` powinno przybierać wartości z przedziału od 1 do 10. Zgodnie z pakietem *mlr3tuningspaces* (Becker, 2023), hiperparametr `num.trees` powinien być ustawiony w zakresie od 1 do 2000, jednak ograniczono jego wartości do przedziału od 50 do 500.

Optymalne hiperparametry uzyskane w wyniku dostrajania modeli lasów losowych dla poszczególnych wariantów (zbiorów danych) razem z oceną jakości AUC zostały przedstawione w tabeli 4.2. Warto zauważyć, że warianty 3 i 6 wykazują identyczne wartości dla parametrów `sample.fraction`, `min.node.size`, `num.trees` i `max.depth`. Podczas losowania hiperparametrów dla każdego wariantu wykorzystano to samo ziarno losowości (ang. *random seed*), ustawione za pomocą funkcji `set.seed()`. Dla dwóch wspomnianych wariantów najlepsze wyniki zostały osiągnięte przy wykorzystaniu identycznego zestawu czterech z pięciu optymalizowanych hiperparametrów.

4.4.3 Ocena jakości modeli

Ważnym krokiem w zastosowaniach uczenia maszynowego jest ocena jakości modelu w rozważanym zadaniu. W tym celu można zastosować k -krotną walidację krzyżową (ang. *k-fold cross-validation*, CV), która zakłada, że obserwacje są od siebie niezależne (Pohjankukka et al., 2017). W standardowej k -krotnej walidacji krzyżowej dostępny zbiór uczący jest dzielony na k podzbiorów podobnej wielkości, gdzie *fold* odnosi się do liczby powstały podzbiorów. Podział ten przeprowadza się poprzez losowe próbkowanie obserwacji ze zbioru uczącego się bez zastępowania. Model jest uczony na $k - 1$ podzbiorach, które razem tworzą zbiór uczący. Następnie model jest testowany na pozostałym podzbiorze, określonym jako zbiór walidacyjny i mierzoną jest jego jakość. Procedurę tę powtarza się, aż każdy z k podzbiorów zostanie użyty jako zbiór walidacyjny (Berrar, 2019). Kroswalidacja pozwala na ocenę ogólnej wydajności modelu, uwzględniając różnorodność danych i pomaga uniknąć sytuacji, w której wyniki są mocno uzależnione od konkretnego podziału danych.

Obserwacje geograficzne, posiadające określone współrzędne, nie spełniają założenia niezależności danych ze względu na autokorelację przestrzenną (ang. *spatial autocorrelation*, SAC) (Pohjankukka et al., 2017). Ogólnie rzecz biorąc, dane przestrzenne wykazują autokorelację przestrzenną zgodnie z pierwszym prawem geografii, będącym jednocześnie podstawowym założeniem analizy geostatystycznej, według którego „Wszystko jest powiązane ze wszystkim innym, ale rzeczy bliskie są bardziej powiązane niż rzeczy odległe” (Tobler, 1970). Traktowanie zbiorów danych przestrzennych jak



Rycina 4.4: Porównanie przestrzennego i losowego podziału zbioru danych na potrzeby walidacji krzyżowej jednego powtórzenia. Podział przestrzenny (górny rząd) i losowy (dolny rząd). Opracowanie własne na podstawie Schratz, 2024

nieprzestrzennych prowadzi do zbyt optymistycznych wyników w zakresie dokładności predykcyjnej modeli (Brenning, 2005) Konsekwencją autokorelacji przestrzennej dla oceny jakości jest nadmierne dopasowanie klasyfikatorów do obserwacji uczących jeśli obserwacje testowe (lub walidacyjne) nie są niezależne od zbioru uczącego (Brenning, 2012). Przestrzenna walidacja krzyżowa (ang. *spatial cross-validation*) jest modyfikacją standardowej kroswalidacji, która zapobiega błędom w ocenie jakości modelu wynikającym z bliskości danych testowych i treningowych. Aby ograniczyć stronniczość w wynikach oceny dokładności predykcyjnej, w ramach przestrzennej walidacji krzyżowej wykorzystuje się przestrzennie odseparowane podzbiory danych, wprowadzając przestrzenną odległość pomiędzy zbiorem treningowym a testowym (Pohjankukka et al., 2017). Przykład takiego podejścia przedstawia rycina 4.4.

Ocenę klasyfikatorów stworzonych dla każdego z wariantów (tabela 4.1) przeprowadzono przy pomocy pięciu miar jakości, przeznaczonych dla klasyfikatorów binarnych, wykorzystując przestrzenną walidację krzyżową:

- precyza (ang. *precision*, inaczej *positive predictive value*) - określa jaką część wyników wskazanych przez klasyfikator jako dodatnie jest faktycznie dodatnia (Jaworski, 2013)
- czułość (ang. *sensitivity*, inaczej *recall* lub *true positive rate*) - określa jaką część dodatnich wyników wykrył klasyfikator (Jaworski, 2013)
- specyficzność (ang. *specificity*, inaczej *true negative rate*) - określa jaką część ujemnych wyników wykrył klasyfikator (Jaworski, 2013)
- pole powierzchni pod krzywą (ang. *Area Under Curve* lub *Area Under the ROC Curve*, AUC lub AUROC) - oblicza obszar pod krzywą ROC, która graficznie przedstawia zależność pomiędzy czułością a specyficznością (Jaworski, 2013). AUC ocenia prawdopodobieństwo, że losowo wybrana obserwacja pozytywna ma wyższe przewidywane prawdopodobieństwo niż losowo wybrana jednostka negatywna (Lang, 2022)
- F_β score (F-beta score) - ważona średnia harmoniczna pomiędzy precyzją i czułością, umożliwiająca ocenę balansu między czułością a precyzją, która w pewnym stopniu opisuje całościowo wynik. Miara ta nie uwzględnia wyników prawdziwie negatywnych (Żygierewicz, 2021)

Wartości każdej z wymienionych miar jakości modelu zawierają się w zakresie od 0 do 1, gdzie wartość 0 reprezentuje niską jakość modelu, natomiast wartość 1 odzwierciedla wysoki poziom jakości. Wartości AUC mieszczą się w zakresie od 0 do 1, gdzie wartość 0,5 lub niższa oznacza model nie lepszy od losowego, a 1,0 – doskonałe przewidywanie obu klas. Wyższa wartość AUC wskazuje na lepszą moc predykcyjną modelu (Lovelace et al., 2019).

4.5 Ważność zmiennych

Ocena ważności zmiennych (ang. *variable importance*) może zostać wykorzystana do poprawy i oceny jakości stworzonych modeli. Analizowanie wpływu poszczególnych zmiennych na dokładność modelu umożliwia ocenę ich istotności dla przewidywań.

Stosowanie zmiennych o niskiej mocy predykcyjnej może prowadzić do nadmiernego dopasowania (ang. *overfitting*) modelu lub obniżenia jego jakości. Dlatego ważny jest wybór odpowiednich zmiennych do trenowania modelu, unikając na przykład kolinearności predyktorów, czyli wysokiej korelacji między zmiennymi. Celem określania ważności zmiennych i ich selekcji jest zwiększenie mocy predykcyjnej modelu w kontekście analizowanego zjawiska poprzez identyfikację silnie skorelowanych z nim zmiennych.

W lasach losowych ważność zmiennych można ocenić różnymi metodami, z których dwie najpopularniejsze to miara zanieczyszczenia Giniego (ang. *Gini impurity*) oraz metoda oparta na permutacji (Biecek, 2016). W celu określenia wartości zmiennych wykorzystywanych do identyfikacji farm fotowoltaicznych na podstawie danych teledetekcyjnych, zastosowano metodę permutacji (ang. *permutation*), która może być również używana do upraszczania i eksploracji modeli lub generowania wiedzy (Biecek et al., 2021).

Główną ideą metody opartej na permutacji jest pomiar tego, jak bardzo zmieni się dopasowanie modelu, gdy usunięty zostanie wpływ wybranej zmiennej lub grupy zmiennych. Jeśli zmienna jest istotna, permutacja jej wartości skutkuje pogorszeniem jakości modelu. Im większa zmiana dopasowania modelu, tym istotniejsza jest permutowana zmienna (Biecek et al., 2021).

Metoda oparta na permutacji została pierwotnie zaproponowana przez Breiman (2001) dla lasów losowych, jednak jej prostota umożliwia zastosowanie permutacji do dowolnego modelu, a także porównywanie ważności zmiennych pomiędzy modelami o różnych strukturach (Biecek et al., 2021).

4.6 Oprogramowanie

4.6.1 QGIS

QGIS (QGIS Development Team, 2023), to wieloplatformowe i wolne oprogramowanie o otwartym kodzie źródłowym przeznaczone do przetwarzania danych przestrzennych, rozwijane 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 zobrazowań satelitarnych (Hejmanowska et al., 2020).

Oprogramowanie QGIS zostało wykorzystane w tej pracy 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.

4.6.2 Sentinel-1 Toolbox i SNAP

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), korekcji, ortorektyfikacji, mozaikowania, konwersji danych, polarymetrii czy interferometrii (ESA, 2023i). Sentinel-1 Toolbox jest opracowywany dla ESA przez firmę Array we współpracy z DLR, Brockmann Consult i OceanDataLab (ESA, 2023i).

SNAP (ESA, 2023n), 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, 2023o,m). Oprogramowanie SNAP zawiera zestawy narzędzi do wizualizacji, przetwarzania oraz analizy danych teledetekcyjnych, a także umożliwia tworzenie łańcuchów procesów przetwarzania danych zdefiniowanych przez użytkownika (Hejmanowska et al., 2020; Moskolaï et al., 2022).

4.6.3 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. Dzięki dużej liczbie pakietów R obsługuje również statystki 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 *GLCMTexels* (Ilich, 2023). Losowe generowanie danych przestrzennych umożliwił 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., 2023a), w ramach którego użyty został algorytm lasów losowych zaimplementowany w pakiecie *ranger* (Wright et al., 2023). Podczas strojenia parametrów algorytmów uczenia maszynowego, korzystano również z pakietu *paradox* (Lang et al., 2023b), umożliwiającego definiowanie granic przestrzeni szukania optymalnych wartości hiperparametrów. 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.

Rozdział 5

Wyniki

5.1 Ocena jakości modeli

5.2 Ważność zmiennych

5.3 Wizualna kontrola wyników klasyfikacji

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 przeходить od szczegółu do ogólnego - 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

- Attema, E, P Snoeij, M Davidson, N Flouri, G Levrini, B Rommen, and B Rosich (2008). The European GMES Sentinel-1 Radar Mission. In: vol. 1, pp.I–94.
- 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 3.2-2. <http://spatstat.org/>.
- Becker, M (2023). *mlr3tuningspaces: Search Spaces for mlr3*. R package version 0.4.0. <https://CRAN.R-project.org/package=mlr3tuningspaces>.
- 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 1.33.1. <https://CRAN.R-project.org/package=future>.
- Bennett, J (2010). *OpenStreetMap*. Packt Publishing Ltd.
- Berrar, D (2019). “Cross-Validation”. In: *Encyclopedia of Bioinformatics and Computational Biology*. Ed. by S Ranganathan, M Gribkov, K Nakai, and C Schönbach. Oxford: Academic Press, pp.542–545. <https://www.sciencedirect.com/science/article/pii/B978012809633820349X>.
- Biecek, P (2016). *Przewodnik: Datasets and Functions Used in the Book Przewodnik po Pakiecie R*. R package version 0.16.12.
- Biecek, P and T Burzykowski (2021). *Explanatory Model Analysis*. Chapman and Hall/CRC, New York. <https://pbiecek.github.io/ema/>.
- Bischl, B, R Sonabend, L Kotthoff, and M Lang, eds. (2024). *Applied Machine Learning Using mlr3 in R*. CRC Press. <https://mlr3book.mlr-org.com>.

- Braun, A (2020). *SAR-based Landcover Classification with Sentinel-1 GRD Products*. SkyWatch Space Applications Inc. <http://step.esa.int>.
- Breiman, L (2001). Random Forests. *Machine Learning* **45**, 5–32.
- Brenning, A (2005). Spatial prediction models for landslide hazards: review, comparison and evaluation. *Natural Hazards and Earth System Sciences* **5**(6), 853–862.
- Brenning, A (2012). Spatial cross-validation and bootstrap for the assessment of prediction rules in remote sensing: The R package sperrorest. In: *2012 IEEE International Geoscience and Remote Sensing Symposium*, pp.5372–5375.
- 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*. <https://sentinels.copernicus.eu/web/sentinel/user-guides/document-library/-/asset%5C%5Fpublisher/xlslt4309D5h/content/sentinel-2-user-handbook>.
- European Space Agency (2023). *Copernicus: Sentinel-1*. Accessed on 2023-12-12. <https://www.eoportal.org/satellite-missions/copernicus-sentinel-1#performance-specifications>.
- European Space Agency (5, 2023a). *ESA Sentinel-1 Toolbox*. Version 9.0.4. <http://step.esa.int/>.
- European Space Agency (2023b). *Mission ends for Copernicus Sentinel-1B satellite*. Accessed on 2023-12-12. https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-1/Mission_ends_for_Copernicus_Sentinel-1B_satellite.
- European Space Agency (2023c). *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 (2023d). *Sentinel-1 Data Products*. Accessed on 2023-12-08. <https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-1/data-products>.

- European Space Agency (2023e). *Sentinel-1 Instrument Payload*. Accessed on 2023-12-08.
<https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-1/instrument-payload>.
- European Space Agency (2023f). *Sentinel-1 Instrument Payload: Resolution and Swath*. Accessed on 2023-12-18. <https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-1/instrument-payload/resolution-swath>.
- European Space Agency (2023g). *Sentinel-1 Mission Objectives*. Accessed on 2023-12-08.
<https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-1/mission-objectives>.
- European Space Agency (2023h). *Sentinel-1 Revisit and Coverage*. Accessed on 2023-12-12. <https://sentinel.esa.int/web/sentinel/user-guides/sentinel-1-sar/revisit-and-coverage>.
- European Space Agency (5, 2023i). *Sentinel-1 Toolbox*. Accessed on 2023-12-05. <https://sentinel.esa.int/web/sentinel/toolboxes/sentinel-1>.
- European Space Agency (2023j). *Sentinel-1 User Guide: Definitions*. Accessed on 2023-12-18. <https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-1-sar/definitions>.
- European Space Agency (2023k). *Sentinel-2 Mission Guide*. Accessed on 2023-12-14. <https://sentinel.esa.int/web/sentinel/missions/sentinel-2>.
- European Space Agency (2023l). *Sentinel-2 User Guide: Spectral Resolution*. Accessed on 2023-12-18. <https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi/resolutions/spectral>.
- European Space Agency (2023m). *SNAP*. Accessed on 2023-12-05. <https://step.esa.int/main/toolboxes/snap/>.
- European Space Agency (5, 2023n). *SNAP - ESA Sentinel Application Platform*. Version 9.0.8. <http://step.esa.int>.
- European Space Agency (5, 2023o). *SNAP Desktop*. Accessed on 2023-12-05. <https://github.com/senbox-org/snap-desktop>.
- 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 (2023). *Ortofotomapa*. Accessed on 2023-12-09. <https://www.geoportal.gov.pl/pl/dane/ortofotomapa-ortho/>.
- 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>.
- Hanssen, RF (2001). *Radar Interferometry Data Interpretation and Error Analysis*. Vol. 2. Springer Science & Business Media.
- 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 1.7-65. <https://rspatial.org/>.
- Ilich, A (2023). *GLCMTexels: GLCM Textures of Raster Layers*. R package version 0.3.9, <https://github.com/ailich/GLCMTexels>. <https://ailich.github.io/GLCMTexels/>.
- 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*.
- Jaworski, W (2013). *Miary jakości*. https://www.mimuw.edu.pl/~wjaworski/SU/SU04_miary_jakosci.pdf.
- 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 (2022). *mlr3measures: Performance Measures for mlr3*. R package version 0.5.0. <https://CRAN.R-project.org/package=mlr3measures>.
- Lang, M, B Bischl, J Richter, P Schratz, M Binder, F Pfisterer, R Sonabend, and M Becker (2023a). *mlr3: Machine Learning in R - Next Generation*. R package version 0.17.1, <https://github.com/mlr-org/mlr3>. <https://mlr3.mlro.org.com>.
- Lang, M, B Bischl, J Richter, X Sun, and M Binder (2023b). *paradox: Define and Work with Parameter Spaces for Complex Algorithms*. R package version 0.11.1, <https://github.com/mlr-org/paradox>. <https://paradox.mlro.org.com>.
- Lesiv, M, L See, JC Laso Bayas, T Sturn, D Schepaschenko, M Karner, I Moorthy, I McCalum, 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 tekstuury 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>.
- 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 (2023). *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>.
- Moskolaï, 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.

BIBLIOGRAFIA

- 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.
- Nowosad, J (2021). *Geostatystyka w R*. Space A. <https://doi.org/10.5281/zenodo.3613745>.
- OpenStreetMap contributors (2024). *Planet dump retrieved from https://planet.osm.org*. <https://www.openstreetmap.org>.
- Pebesma, E (2023). *sf: Simple Features for R*. R package version 1.0-15, <https://github.com/r-spatial/sf>. <https://r-spatial.github.io/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. <https://assets.planet.com/products/basemap/planet-basemaps-product-specifications.pdf>.
- Pohjankukka, J, T Pahikkala, P Nevalainen, and J Heikkonen (2017). Estimating the prediction performance of spatial models via spatial k-fold cross validation. *International Journal of Geographical Information Science* **31**(10), 2001–2019. eprint: <https://doi.org/10.1080/13658816.2017.1346255>.
- Probst, P, MN Wright, and AL Boulesteix (2019). Hyperparameters and tuning strategies for random forest. *WIREs Data Mining and Knowledge Discovery* **9**(3), e1301. eprint: <https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/widm.1301>.
- 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*.
- Schratz, P, J Muenchow, E Iturritxa, J Richter, and A Brenning (2019). Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data. *Ecological Modelling* **406**, 109–120.
- Sekulić, A, M Kilibarda, GB Heuvelink, M Nikolić, and B Bajat (2020). Random Forest Spatial Interpolation. *Remote Sensing* **12**(10).
- Sheykhmousa, 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*.
- Tobler, WR (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography* **46**, 234–240. (Visited on 01/05/2024).
- 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).
- Wright, MN, S Wager, and P Probst (2023). *ranger: A Fast Implementation of Random Forests*. R package version 0.16.0, <https://github.com/imbs-hl/ranger>. <http://imbs-hl.github.io/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).

BIBLIOGRAFIA

Żygierewicz, J (2021). *Uczenie maszynowe i sztuczne sieci neuronowe: Ocena jakości klasyfikacji.*

Accessed on 2024-01-04. https://brain.fuw.edu/index.php/Uczenie_maszynowe_i_sztuczne_sieci_neuronowe/Wyk%C5%82ad_Ocena_jako%C5%9Bci_klasyfikacji.