



# Pemetaan Otomatis Air Permukaan dengan Pendekatan Random Forest dan Citra Sentinel-2

Muh Farhan<sup>1,2</sup>, Setia Pramana<sup>2</sup>

<sup>1</sup>Program Studi D-IV Komputasi Statistik, Politeknik Statistika STIS

<sup>2</sup>Badan Pusat Statistik

<sup>1</sup>[222112195@stis.ac.id](mailto:222112195@stis.ac.id)

<sup>2</sup>[setia.pramana@stis.ac.id](mailto:setia.pramana@stis.ac.id)

Corresponding author email: [222112195@stis.ac.id](mailto:222112195@stis.ac.id)

**Abstract:** Understanding the presence and distribution of the Earth's surface water is crucial for various environmental planning segments such as water resources management, natural disaster mitigation, environmental conservation, and spatial planning. In addition, the availability of Sentinel-2 satellite imagery data that offers high spatial resolution with wide coverage allows for more accurate and detailed surface water mapping compared to traditional methods. This research presents the utilization of the Random Forest algorithm as a surface water mapping model tested on Sentinel-2 image data and can be applied nationally with a spatial resolution of 10 meters. In this case, the Random Forest model was developed using spectral indices such as Modified Normalized Difference Water Index (MNDWI) and Normalized Vegetation Index (NDVI) as well as true color (RGB) spectral bands. The accuracy assessment results show that the Random Forest model built can perform surface water mapping very accurately, which is confirmed by the evaluation metrics in the form of overall accuracy, average producer's accuracy, and average user's accuracy which have a value of 98.12%, 98.13%, and 98.07%, respectively. Based on these results, it can be seen that the model has significant potential to assist the environmental planning process further through surface water mapping.

**Keywords:** Surface water, Automatic mapping, Sentinel-2, Random Forest

**Abstrak:** Memahami keberadaan dan distribusi air permukaan bumi merupakan hal krusial bagi berbagai segmen perencanaan lingkungan seperti pengelolaan sumber daya air, mitigasi bencana alam, pelestarian lingkungan, hingga perencanaan tata ruang. Di samping itu, keberadaan data citra satelit Sentinel-2 yang menawarkan resolusi spasial tinggi dengan cakupan yang luas memungkinkan pemetaan air permukaan yang lebih akurat dan detail dibandingkan dengan metode tradisional. Penelitian ini menyajikan pemanfaatan algoritma Random Forest sebagai model pemetaan air permukaan yang diuji pada data citra Sentinel-2 dan dapat diaplikasikan secara nasional dengan resolusi spasial 10 meter. Dalam hal ini model Random Forest dikembangkan menggunakan indeks spectral berupa Modified Normalized Difference Water Index (MNDWI) dan Normalized Vegetation Index (NDVI) serta band spectral true color (RGB). Hasil penilaian akurasi menunjukkan bahwa model Random Forest yang dibangun dapat melakukan pemetaan air permukaan dengan sangat akurat yang dikonfirmasi oleh metrik evaluasi berupa overall accuracy, rata-rata producer's accuracy, dan rata-rata user's accuracy yang secara berturut-turut memiliki nilai 98,12%, 98,13%, dan 98,07%. Berdasarkan hasil tersebut, terlihat bahwa model tersebut memiliki potensi yang signifikan untuk membantu proses perencanaan lingkungan lebih dalam melalui pemetaan air permukaan.

**Kata kunci:** Air permukaan, Pemetaan otomatis, Sentinel-2, Random Forest

## I. PENDAHULUAN

Air memainkan peran penting dalam kelangsungan hidup banyak makhluk hidup. Air tersedia dalam berbagai bentuk termasuk air tanah, badan air permukaan, dan badan air yang membeku, dan sebagainya. Oleh karena itu, penting untuk menjaga dan melestarikan sumber daya alam ini karena ekosistem sangat bergantung padanya. Kekurangan air dapat menyebabkan banyak masalah baik bagi manusia maupun organisme lainnya. Sumber daya air ini membutuhkan pengelolaan yang efektif untuk mencegah terjadinya korban lingkungan [1]. Tercatat sekitar 27% atau 2,2 miliar orang di dunia belum memiliki akses layanan air minum yang aman pada tahun 2022. Sedangkan jumlah orang yang belum memiliki akses layanan sanitasi yang aman mencapai 43% atau 3,5 miliar orang pada tahun 2022 [2]. Oleh karena itu, ada kebutuhan untuk penilaian massal terhadap sumber daya air permukaan dengan menggunakan teknologi yang efisien dan layak [1].



Dengan pesatnya perkembangan teknologi kedirgantaraan, teknologi penginderaan jauh menyediakan metode-metode canggih untuk berbagai bidang seperti survei sumber daya, pemantauan lingkungan, analisis regional, dan penelitian makro global [3]. Penginderaan jauh sendiri merujuk pada teknik-teknik yang dikembangkan untuk mengambil atau memperoleh informasi dan analisis mengenai permukaan bumi. Salah satu hasil dari beragam hasil proses kegiatan penginderaan jauh adalah citra yang berupa gambar objek atau area yang berada pada permukaan bumi yang ditangkap menggunakan lensa kamera. Gambar ini diperoleh dengan menangkap gelombang yang dipantulkan maupun dihasilkan oleh permukaan bumi untuk kemudian dijadikan sebagai input dalam metode-metode analisis penginderaan jauh lanjutan [4]. Dalam lingkup pemetaan, Salwa dkk. berhasil memanfaatkan citra satelit multisumber dalam rangka memetakan kemiskinan di Provinsi Jawa Tengah yang salah satunya memanfaatkan *Normalized Difference Water Index* (NDVI) sebagai pendekatan dalam mendeteksi tutupan lahan terutama untuk membedakan daerah perairan dan daratan karena nilainya yang sensitif terhadap kadar air [5]. Ekstraksi informasi air yang akurat dari citra penginderaan jauh sangat penting untuk perencanaan dan investigasi daerah aliran sungai, pemantauan, dan perlindungan sumber daya air [6]. Pada satu sisi, Indonesia dalam satu dekade terakhir telah muncul sebagai pusat maritim yang menghubungkan rute pelayaran Asia-Afrika-Eropa dan menjadi penting pada aspek ekonomi dan geopolitik negara ini. Lingkungan perairan negara ini juga relatif kompleks karena mencakup banyak jenis badan air, seperti danau, waduk, kolam, laguna, dan sungai. Indonesia merupakan wilayah tropis, sehingga citra satelit pengamatan bumi dipengaruhi oleh tutupan awan dan bayangan awan sepanjang tahun. Pada saat yang sama, terdapat banyak sawah yang tertanam dengan padi di pulau ini, dan karakteristik spektral air mendekati karakteristik tanah setelah panen padi, terutama setelah hujan. Faktor-faktor ini meningkatkan kesulitan untuk mendapatkan informasi air permukaan yang akurat dari citra penginderaan jauh.

Dalam beberapa tahun terakhir, para ahli di seluruh dunia telah melakukan penelitian ekstensif tentang ekstraksi air dari citra berbagai sensor satelit. Berdasarkan hubungan antara *band* kedua dan keempat dari citra TM, Mcfeeters dkk. mengusulkan *normalized difference water index* (NDWI), yang secara efektif dapat menekan informasi tanah dan vegetasi, tetapi sulit untuk menekan informasi bangunan [7]. Zhou dkk. menemukan bahwa badan air pada citra TM memiliki karakteristik hubungan spektral yang unik, yaitu  $(TM2+TM3) > (TM4+TM5)$ , yang sangat cocok untuk ekstraksi air di daerah pegunungan. Berdasarkan NDWI, Xu memperkenalkan *modified normalized difference water index* (MNDWI), yang menunjukkan akurasi yang lebih tinggi, terutama pada ekstraksi air di daerah perkotaan [8]. Yan dkk. mengusulkan indeks air yang disempurnakan (*enhanced water index*, EWI), dan dikombinasikan dengan teknologi GIS untuk mengekstraksi air permukaan secara efektif di daerah semi-kering [9]. Berdasarkan citra Landsat-8 OLI, Zhang dkk. membangun model ekstraksi air berdasarkan transformasi LBV yang memiliki akurasi lebih baik dibandingkan NDWI dan MNDWI [10]. Menggabungkan citra Sentinel-2 dan data OpenStreetMap, Zhang dkk. mengusulkan model ekstraksi badan air berdasarkan algoritma pembelajaran keberadaan dan latar belakang, dan membuktikan bahwa model tersebut memiliki akurasi yang lebih tinggi daripada eksponensial badan air dan hutan acak melalui eksperimen [11]. Para ahli di seluruh dunia telah mengusulkan banyak model ekstraksi air untuk sensor yang berbeda, terutama berfokus pada lingkungan konvensional, ekstraksi air berskala besar dan berefisiensi tinggi [12], namun relatif jarang memperhatikan ekstraksi air yang akurat di lingkungan yang kompleks. Misalnya, di daerah berawan, pegunungan, dan tanah sawah seperti Indonesia, model konvensional relatif tidak dapat secara efektif menghilangkan pengaruh



bayangan awan, bayangan gunung, dan tanah sawah terhadap keakuratan ekstraksi air, sehingga ekstraksi air permukaan yang akurat dihadapkan pada tantangan.

Satelit Sentinel-2 merupakan satelit kedua dalam program Copernicus dari Komisi Eropa (European Commission) dan Badan Antariksa Eropa (European Space Agency (ESA) [13]. Tugas utama satelit Sentinel-2 adalah mewujudkan pencitraan multispektral resolusi permukaan tanah global. Dibandingkan dengan citra Landsat-8, citra Sentinel-2 memiliki resolusi spasial dan temporal yang lebih tinggi. Dalam hal ekstraksi air menggunakan indeks badan air (NDWI, MNDWI), akurasi citra Sentinel-2 lebih baik dibandingkan citra Landsat-8, terutama dalam ekstraksi sungai [14]. Citra Sentinel-2 memiliki *band* biru, hijau, merah, dan *near-infrared* dengan resolusi 10 m. Wang dkk. memilih empat *band* beresolusi 10 m dan dua band inframerah gelombang pendek beresolusi 20m dari citra Sentinel-2 sebagai fitur input, dan dilatih melalui SVM untuk menemukan hyperplane pemisahan yang optimal, dan mengusulkan Indeks Air Multispektral (*Multispectral Water Index/MuWI*), yang menghasilkan pemetaan sumber daya air permukaan dengan resolusi 10m [15]. Namun, MuWI sulit untuk digunakan secara universal karena keterbatasan ukuran sampel pelatihan. Untuk mencapai ekstraksi air dengan presisi tinggi di lingkungan yang kompleks, model ekstraksi air permukaan menggunakan model *Random Forest Classification* dibangun berdasarkan *platform* komputasi awan penginderaan jauh Google Earth Engine dan citra Sentinel-2 pada keseluruhan wilayah Indonesia dengan tujuan untuk memetakan dan memprediksi keberadaan air permukaan di wilayah Indonesia sebagai upaya untuk menyediakan informasi yang akurat kepada pihak berwenang yang kemudian dapat digunakan secara adaptif dalam rangka mengembangkan kebijakan-kebijakan penngelolaan lingkungan khususnya pada domain perairan.

## II. METODE PENELITIAN

### 2.1. Area Studi

Pembangunan model pemetaan air permukaan dilakukan di seluruh wilayah Indonesia yang secara astronomis berada di antara 6° Lintang Utara dan 11° Lintang Selatan serta 95° Bujur Timur dan 141° Bujur Timur. Negara ini memiliki sekitar 17,500 pulau dengan luas wilayah daratan sebesar 1.91 juta km persegi dan luas wilayah perairannya mencapai 6.32 juta km persegi. Secara umum, distribusi air permukaan di Indonesia menurut Menteri PUPR Basuki Hadimuljono tidak merata. Sebanyak 82% di antaranya berada di Pulau Kalimantan, Papua, dan Sumatera. Sedangkan Jawa yang penggunaannya tertinggi hanya memiliki 6,3% atau setara dengan 175,6 miliar m kubik per tahun [16]. Berdasarkan karakteristik distribusi air permukaan tersebut, maka titik-titik sampel untuk proses *training* model diambil secara acak pada pulau-pulau besar di Indonesia yang dalam hal ini adalah Pulau Sumatera, Pulau Jawa, Pulau Kalimantan, Pulau Sulawesi, Kepulauan Maluku, dan Pulau Papua. Sebaran titik-titik sampel yang digunakan disajikan pada gambar 1 berikut ini.



**Gambar 1.** Distribusi Titik Sampel dengan ■ Merupakan Area Non-air dan ■ Merupakan Area Air (Masing-masing Jenis Area Memiliki 1000 Titik Sampel)



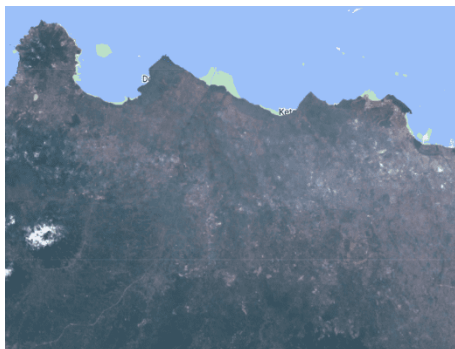
## 2.2. Data Akuisisi dan Preprocessing

### 2.2.1. Large Scale International Boundary (LSIB) Polygons, Simplified

Large Scale International Boundary (LSIB) menyediakan data batas internasional dengan presisi tinggi dalam skala global, termasuk informasi mengenai batas-batas di lebih dari 250 negara, wilayah, dan teritori. Dataset ini disediakan oleh United States Office of Geography dan berisi file vektor garis LSIB dan Garis Pantai Vektor Dunia (World Vector Shorelines, WVS) dari Badan Intelijen Geospasial Nasional (National Geospatial-Intelligence Agency, NGA) [17]. Dalam penelitian ini, LSIB digunakan untuk mengambil poligon wilayah negara Indonesia sebagai batas lingkup kerja pembangunan model.

### 2.2.2. Harmonized Sentinel-2 Multispectral Instrument

Sentinel-2 merupakan satelit pencitraan multispektral resolusi tinggi, yang terdiri dari dua satelit, yaitu 2A dan 2B. Sentinel-2A diluncurkan pada tanggal 23 Juni 2015 dan Sentinel-2B telah berada di orbit sejak 7 Maret 2017, dengan resolusi temporal lima hari di khatulistiwa [4]. Sentinel-2 membawa multi-spektrometer dengan 13 pita spektral yang mencakup cahaya tampak, inframerah dekat, dan inframerah gelombang pendek [18]. Secara khusus, data Sentinel-2 dalam penelitian ini berasal dari salah satu set data Google Earth Engine yaitu “COPERNICUS/S2\_HARMONIZED” yang seusai untuk studi pemantauan vegetasi, identifikasi tutupan tanah dan air, serta pengamatan jalur air pedalaman dan wilayah pesisir [5]. Pada penelitian ini, data citra Sentinel-2 yang digunakan merupakan data citra pada periode 1 Januari 2023 sampai dengan 1 Januari 2024 yang kemudian difilter berdasarkan persentase pixel awan dengan kriteria di bawah 10% agar diperoleh data citra yang relatif bersih dari tutupan awan. Kemudian untuk mendapatkan sebuah citra komposit dari kumpulan citra selama periode 2023, maka dilakukan mekanisme reduksi dengan menggunakan median pixel dari kumpulan citra yang ada. Nilai-nilai *band* dan indeks spectral seperti *Modified Normalized Difference Water Index* (MNDWI) dan *Normalized Vegetation Index* (NDVI) selanjutnya diekstrak dari citra komposit tersebut. Contoh tampilan citra komposit dari Sentinel-2 yang digunakan dalam penelitian ini disajikan pada gambar 2 berikut ini.



Gambar 2a.



Gambar 2b.

**Gambar 2.** Citra Komposit Sentinel-2 dengan *True Color Composite* (band 4, 3, 2): (2a) Wilayah Jakarta (2b) Wilayah Kepulauan Bangka Belitung

## 2.3. Pemilihan Indeks Spektral

Dalam praktiknya, banyak peneliti menggunakan indeks spektral tertentu untuk membuat satu kelas tutupan lahan tertentu lebih menonjol dibandingkan tutupan lahan lainnya. Beragam indeks spektral telah dikembangkan, misalnya untuk menonjolkan vegetasi, *Normalized Difference Vegetation Index* (NDVI) dapat digunakan. Untuk lahan gundul, *Normalized Difference Bareness Index* (NDBaI)





dan *Normalized Difference Bare Land Index* (NBLI) dapat digunakan. Untuk ekstraksi air, salah satu antara *Normalized Difference Water Index* (NDWI) dan *Modified Normalized Difference Water Index* (MNDWI) [19]. Khusus dalam penelitian ini digunakan MNDWI untuk mengekstraksi badan air, mengikuti penelitian sebelumnya oleh Acharya et al., 2018 yang menguji MNDWI [20] dan penelitian lain oleh Gasparovich et al., 2019 yang merekomendasikan penggunaan MNDWI untuk pemetaan badan air [21]. Dalam melakukan kalkulasi MNDWI, formula yang digunakan adalah sebagai berikut:

$$MNDWI = \frac{Green - SWIR}{Green + SWIR} \quad (1)$$

Dengan SWIR direpresentasikan oleh *band* 11 dari Sentinel-2 yang memiliki panjang gelombang 1610 nm dan *Green* direpresentasikan oleh *band* 3 Sentinel-2 dengan panjang gelombang 560 nm. Kemudian untuk bisa menangkap lebih presisi keberadaan air permukaan yang mungkin berada di tengah-tengah lingkungan vegetasi seperti sungai, danau dangkal, dan perairan kecil yang berair jernih maka NDVI juga digunakan. Dalam hal ini, area permukaan yang memiliki NDVI rendah dan MNDWI tinggi akan diidentifikasi sebagai area air permukaan. Secara matematis formula yang digunakan untuk menghitung NDVI adalah sebagai berikut:

$$NDVI = \frac{NIR - Red}{NIR + Red} \quad (2)$$

Dengan NIR direpresentasikan oleh *band* 8 dari Sentinel-2 yang memiliki panjang gelombang 835 nm dan *Red* direpresentasikan oleh *band* 4 Sentinel-2 dengan panjang gelombang 665 nm.

#### 2.4. Klasifikasi Random Forest

*Random Forest* (RF) adalah metode klasifikasi berbasis *decision tree*/pohon keputusan yang menghasilkan banyak *tree* dari sampel-sampel *training set* [22]. Dalam hal ini beberapa pohon keputusan dikombinasikan dan bergantung pada nilai vektor acak yang diambil secara independen dan memiliki distribusi yang sama untuk semua pohon. Klasifikasi RF adalah jenis klasifikasi *ensemble*, yang berarti metode ini menggunakan banyak pengklasifikasi, bukan hanya satu. Dalam klasifikasi RF, ratusan *tree* sebagai pengklasifikasi dibangun dan keputusan mereka biasanya digabungkan menggunakan mekanisme *voting* pluralitas atau *decision* mayoritas dari seluruh *tree* yang terbentuk [22]. Ide utamanya adalah bahwa menggabungkan beberapa pengklasifikasi *ensemble* sering kali memberikan hasil yang lebih akurat dibandingkan dengan hanya menggunakan satu pengklasifikasi [23]. Oleh karena itu, klasifikasi RF sering digunakan dalam pengolahan citra penginderaan jauh. Elemen umum dalam prosedur ini adalah bahwa untuk setiap pohon ke- $k$ , vektor acak  $\theta_k$  dihasilkan secara independen dari vektor acak sebelumnya  $\theta_1, \dots, \theta_{k-1}$ , tetapi memiliki distribusi yang sama. Pohon tersebut dibangun menggunakan set pelatihan dan  $\theta_k$ , menghasilkan pengklasifikasi  $h(x, \theta_k)$ , yang mana  $x$  adalah vektor input [24]. Pengklasifikasi terdiri dari beberapa pohon yang dibangun secara sistematis dengan memilih subset dari komponen vektor fitur secara acak, artinya pohon-pohon tersebut dibangun dalam subruang yang dipilih secara acak untuk menjaga akurasi tertinggi dari data pelatihan dan meningkatkan akurasi generalisasi seiring dengan meningkatnya kompleksitas [25]. Melalui proses *bagging*, ketika klasifikasi RF membangun pohon, ia menggunakan pembagian terbaik dari subset acak fitur input atau variabel prediktif di setiap node, bukan menggunakan variabel split terbaik. Meskipun ini dapat mengurangi kekuatan setiap pohon, pemangkasan mengurangi korelasi antar pohon dan kesalahan generalisasi [26]. Karena pohon-pohon dalam pengklasifikasi RF tumbuh tanpa pemangkasan, waktu untuk menghasilkan model tidak meningkat secara signifikan. Kelas yang



diprediksi dari sebuah observasi dihitung berdasarkan suara mayoritas dari pohon-pohon dalam model RF, dan fungsi diskriminasi didefinisikan sebagai berikut:

$$H(x) = \underset{y}{\operatorname{argmax}} \sum_{i=1}^k I(h_i(x, \theta_k) = y) \quad (3)$$

Dengan  $I(\cdot)$  merupakan fungsi indikator,  $h(\cdot)$  merupakan *single decision tree*, dan  $y$  merupakan *output* variabel. Kemudian  $\underset{y}{\operatorname{argmax}}$  menunjukkan nilai  $y$  ketika memaksimalkan  $\sum_{i=1}^k I(h_i(x, \theta_k) = y)$ . Untuk setiap set pelatihan baru yang dibuat untuk membuat pohon tumbuh, sepertiga dari sampel dikeluarkan secara acak, yang disebut *out-of-bag* (OOB). Sampel yang tersisa (*in-the-bag*) digunakan untuk membangun pohon. Sampel OOB ini dapat digunakan untuk mengevaluasi kinerja model, dan telah terbukti bahwa estimasi yang dihasilkan OOB tidak bias [24]. Dalam hal ini, vektor input pada model RF yang digunakan diambil dari nilai-nilai *band true color* (*band 4, band 3, band 2*), nilai *Modified Normalized Difference Water Index* (MNDWI), dan nilai *Normalized Difference Vegetation Index* (NDVI).

### 2.5. Evaluasi Model

Evaluasi model dalam penelitian ini secara khusus difokuskan pada penilaian dan pengujian akurasi yang didasarkan pada perbandingan antara hasil klasifikasi model *Random Forest* dan titik sampel tutupan lahan yang dipilih secara manual berdasarkan kondisi wilayah (wilayah air dan wilayah non-air) pada citra komposit yang diperoleh sebagai *ground truth*. Dalam hal ini sebanyak 2000 titik sampel dibuat dengan 1000 titik sampel untuk wilayah air dan 1000 titik sampel untuk wilayah non-air. Selanjutnya, titik-titik sampel tersebut kemudian dibagi menjadi training set dan test set dengan proporsi masing-masing 80% dan 20%. Seluruh piksel di wilayah studi digunakan untuk penilaian akurasi. Terdapat tiga metrik akurasi yang digunakan untuk dievaluasi yaitu *overall accuracy*, *producer’s accuracy*, dan *user’s accuracy*. *Overall accuracy* pada dasarnya menunjukkan proporsi area yang dipetakan dengan benar dari keseluruhan area yang ada. Sehingga semakin tinggi nilai *overall accuracy*, maka semakin banyak area yang diklasifikasikan dengan benar. Kemudian *producer’s accuracy* adalah akurasi peta dari sudut pandang pembuat peta (produsen). Dalam artian nilai *producer’s accuracy* merepresentasikan jumlah area yang diklasifikasikan secara akurat pada kelas tertentu dibagi dengan jumlah total area untuk kelas tersebut. Sedangkan *user’s accuracy* mengarah pada akurasi dari sudut pandang pengguna peta, bukan pembuat peta. *User’s accuracy* pada dasarnya menunjukkan keandalan sebuah model dengan mempertimbangkan seberapa sering kelas yang ada pada peta akan benar-benar ada di lapangan. Artinya, *user’s accuracy* menghitung jumlah area yang diklasifikasikan dengan benar pada kelas tertentu dibagi dengan jumlah area yang diklasifikasikan pada kelas tersebut [27]. Skema pemetaan klasifikasi dan formula matematis ketiga metrik evaluasi tersebut disajikan sebagai berikut.

**Tabel 1.** Skema pemetaan klasifikasi

Hasil Klasifikasi	Test Data		
	Area Air (A)	Area Non-air (N)	$\Sigma$
Area Air (A)	$N_{AA}$	$N_{AN}$	$\sum AR$
Area Non-air (N)	$N_{NA}$	$N_{NN}$	$\sum NR$
$\Sigma$	$\sum PA$	$\sum PN$	$N$



$$\text{Overall accuracy} = \frac{N_{AA} + N_{NN}}{N} \times 100\% \quad (4)$$

$$\text{Producer accuracy (area air)} = \frac{N_{AA}}{\sum PA} \times 100\% \quad (5)$$

$$\text{Producer accuracy (arena non - air)} = \frac{N_{NN}}{\sum PN} \times 100\% \quad (6)$$

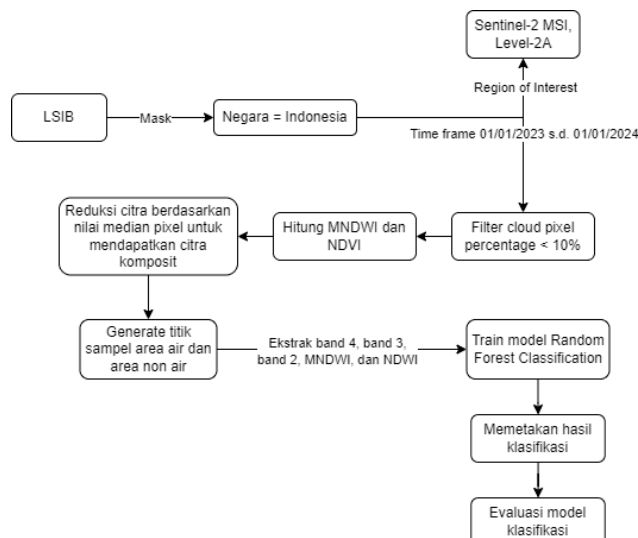
$$\text{User accuracy (arena air)} = \frac{N_{NA}}{\sum AR} \times 100\% \quad (7)$$

$$\text{User accuracy (area non - air)} = \frac{N_{NN}}{\sum NR} \times 100\% \quad (8)$$

Dengan:

- $N_{AA}$  : Jumlah area kelas area air yang berhasil diklasifikasi sebagai kelas area air
- $N_{AN}$  : Jumlah area kelas area non-air yang diklasifikasi sebagai kelas area air
- $N_{NA}$  : Jumlah area kelas area air yang diklasifikasi sebagai kelas area non-air
- $N_{NN}$  : Jumlah area kelas area non-air yang berhasil diklasifikasi sebagai kelas area non-air
- $\sum PA$  : Jumlah area kelas area air pada *testing set*
- $\sum PN$  : Jumlah area kelas area non-air pada *testing set*
- $\sum AR$  : Jumlah area yang diklasifikasi sebagai kelas area air
- $\sum NR$  : Jumlah area yang diklasifikasi sebagai kelas area non-air
- $N$  : Jumlah keseluruhan area yang ada

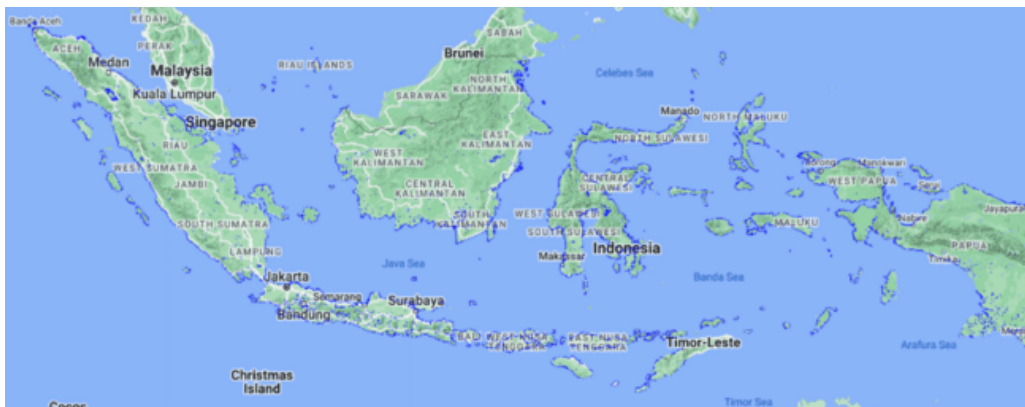
Ketiga metrik evaluasi di atas diterapkan pada *testing set* dengan dua kelas klasifikasi yaitu area air dan area non-air. Kelas-kelas tersebut diekstraksi dari klasifikasi tutupan lahan yang diverifikasi dan dikoreksi secara manual berdasarkan citra *true color composite* dari Sentinel-2. Akhirnya, secara umum, keseluruhan alur metodologi atau tahapan penelitian yang dilakukan digambarkan pada gambar 3 berikut ini.



**Gambar 3.** Flowchart Tahapan Penelitian

### III. HASIL DAN PEMBAHASAN

Untuk menilai akurasi hasil klasifikasi dari model *Random Forest* (RF) yang dibangun, *testing set* yang dipilih dari citra komposit secara acak berdasarkan *true color composite* digunakan sebagai input utama. Nilai-nilai *band 4*, *band 3*, *band 2* sebagai representasi *true color composite* dan nilai MNDWI serta MNDWI dari *testing set* tersebut selanjutnya dijadikan sebagai nilai input pada model RF yang dibangun. Overall *accuracy* yang diperoleh adalah sebesar 98,12%, kemudian *producer’s accuracy* untuk kelas area air yang diperoleh sebesar 98,19% sedangkan untuk kelas area non-air sebesar 98,07%. Selanjutnya nilai *user’s accuracy* untuk kelas area air yang diperoleh adalah sebesar 97,60% sedangkan untuk kelas area non-air sebesar 98,54%. Sehingga, rata-rata nilai *producer’s accuracy* adalah sebesar 98,13% dan rata-rata nilai *user’s accuracy* sebesar 98,07%. Hasil tersebut menunjukkan bahwa klasifikasi dari model RF yang dibangun relatif efektif dalam menangani gangguan latar belakang tanah baik di daerah gersang yang bervegetasi jarang maupun di daerah dengan tutupan vegetasi yang padat. Dengan kata lain, klasifikasi RF yang dibangun memiliki kapasitas genealogi yang cukup kuat dalam proses pelatihan. Hal ini sekaligus membuktikan bahwa strategi *Random Forest Classifier* (RFC) dalam pengambilan keputusan berdasarkan voting gabungan cukup efektif dalam mengidentifikasi area yang termasuk sebagai air permukaan. Selanjutnya, hasil pemetaan area air permukaan berdasarkan klasifikasi dengan model RF disajikan pada gambar 3 berikut ini.



**Gambar 4.** Pemetaan Air Permukaan Berdasarkan Klasifikasi Model *Random Forest*

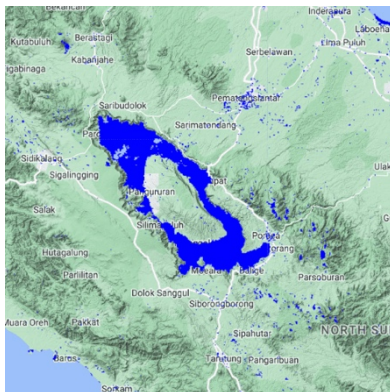
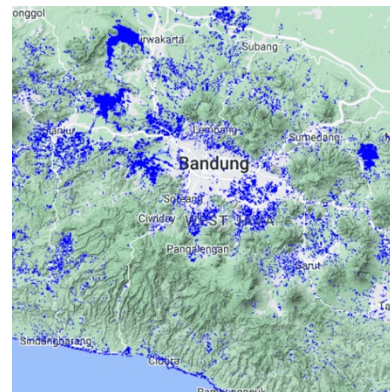
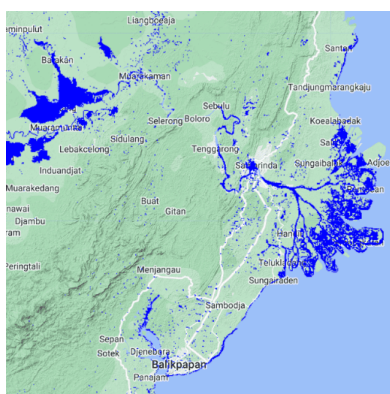
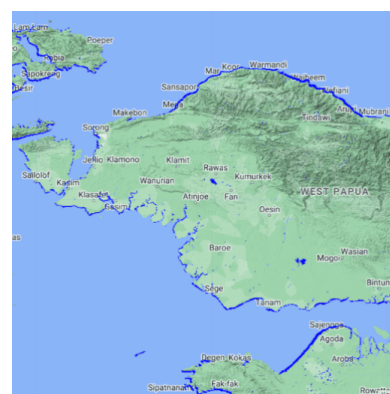
Pada gambar 4 di atas, area yang ditandai dengan warna biru merupakan area yang diidentifikasi sebagai air permukaan. Selanjutnya untuk memberikan perspektif yang lebih detail, gambar 5 berikut ini menyajikan sampel hasil pemetaan air permukaan untuk beberapa wilayah dengan visual yang lebih rinci.

Klasifikasi RF menggunakan subset fitur untuk membangun setiap pohon keputusan. Keuntungan dari pendekatan ini adalah setiap pohon keputusan dapat menghasilkan keputusan klasifikasi yang lebih akurat berdasarkan fitur yang relevan, sehingga mengurangi kesalahan yang timbul dari penggunaan seluruh vektor fitur dalam analisis. Dengan kata lain, setiap pohon keputusan bertindak sebagai spesialis klasifikasi dalam domain fitur yang digunakannya. Untuk fitur yang digunakan dalam membedakan jenis kelas, pohon keputusan non-spesialis melakukan pemungutan suara atau *voting* yang hampir acak. Namun, jumlah pohon keputusan spesialis yang cukup dapat menghasilkan klasifikasi yang akurat ketika jumlah pohon keputusan mencapai batas tertentu. Oleh karena itu, RF memiliki kelebihan dalam hal stabilitas dan ketahanan yang lebih baik.

Indeks spektral memiliki kemampuan untuk mengekstrak fitur, tetapi mereka juga memiliki nilai yang sama untuk kelas tutupan lahan yang berbeda atau tumpang tindih dengan kelas tutupan lahan



lainnya [28]. Dalam hal ini indeks spektral secara khusus digunakan untuk memisahkan kelas perairan dan non-perairan berdasarkan ketidaksamaan spektral. *Modified Normalized Difference Water Index* (MNDWI) berperan sebagai alat yang penting untuk pemantauan lingkungan dan secara efisien membantu menggambarkan wilayah perairan dengan menguraikan air dari wilayah non-air. Selanjutnya indeks spektral *Normalized Difference Vegetation Index* (NDVI) dan pada penelitian ini digunakan untuk mengekstraksi karakteristik vegetasi dari sistem perairan dan mengembangkan model indeks risiko sedimentasi sehingga masalah campuran pixel air dan vegetasi dapat dikurangi. Analisis yang dinormalisasi dari citra MNDWI dan peta air permukaan akhir menunjukkan keefektifan pendekatan berbasis MNDWI dan NDVI [29]. Sehingga, metode yang dibangun pada penelitian ini akan membantu dalam mengembangkan kebijakan praktis untuk memahami ketersediaan sumber daya air di daerah aliran sungai atau danau selama musim kemarau dan kondisi ekstrem lainnya. Pengetahuan ini sangat penting untuk tujuan pengembangan lingkungan yang berkelanjutan di daerah perkotaan, serta adaptasi perubahan iklim perkotaan [30].

**Gambar 5a.****Gambar 5b.****Gambar 5c.****Gambar 5d.**

**Gambar 5.** Pemetaan Air Permukaan: (5a) Wilayah Sumatera Utara (5b) Wilayah Jawa Barat (5c) Wilayah Kalimantan Timur (5d) Wilayah Papua Barat

#### IV. KESIMPULAN

Data satelit yang dapat diakses secara terbuka dapat membantu dalam pemetaan dan pemantauan air permukaan. Dalam penelitian ini, citra multispektral Sentinel-2 digunakan untuk wilayah Indonesia secara geografis dalam rangka memetakan badan air permukaan. Citra multispektral Sentinel-2 memiliki keunggulan dalam hal resolusi spasial yang cukup tinggi hingga 10 m dan cakupan yang



berulang dengan periode orbit 5 hari. *True color composite* (band 4, band 3, band 2) dari Sentinel-2 dan *Modified Normalized Difference Water Index* (MNDWI) serta *Normalized Difference Vegetation Index* (NDVI) digunakan sebagai input pada model *Random Forest* (RF) *Classification* untuk mengekstrak peta air permukaan. Model yang dibangun dapat digunakan untuk pemetaan badan air dengan tingkat akurasi tinggi, yang dikonfirmasi oleh tiga metrik evaluasi yaitu *overall accuracy*, rata-rata *producer’s accuracy*, rata-rata *user’s accuracy* dengan nilai secara berturut-turut adalah 98,12%; 98,13%; dan 98,07%. Metode yang dikembangkan menerapkan pendekatan berbasis piksel dan untuk lebih meningkatkan keakuratan hasil klasifikasinya, algoritma serupa dapat dikembangkan berdasarkan pendekatan berbasis objek di masa mendatang. Selain itu, model RF yang dibangun juga memungkinkan pemetaan global yang cepat dan akurat untuk permukaan air perkotaan dan perdesaan. Oleh karena itu, model ini dapat dengan mudah diterapkan dalam penelitian di masa depan yang berkaitan dengan pemantauan badan air dalam jangka waktu yang lebih lama.

### UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Politeknik Statistika STIS yang telah memberikan dukungan dana penelitian dalam menyelesaikan kajian ini.

### REFERENSI

1. U. Bhangale, S. More, T. Shaikh, S. Patil, and N. More, “Analysis of Surface Water Resources Using Sentinel-2 Imagery,” in *Procedia Computer Science*, Elsevier B.V., 2020, pp. 2645–2654. doi: 10.1016/j.procs.2020.04.287.
2. Kementrian PUPR, “Menteri Basuki: Pemenuhan Air Bersih dan Sanitasi Sebagai Prioritas Agenda Global”. Available online: [https://sahabat.pu.go.id/eppid/page/kilas\\_berita/4015/Menteri-Basuki-Pemenuhan-Air-Bersih-dan-Sanitasi-Sebagai-Prioritas-Agenda-Global-](https://sahabat.pu.go.id/eppid/page/kilas_berita/4015/Menteri-Basuki-Pemenuhan-Air-Bersih-dan-Sanitasi-Sebagai-Prioritas-Agenda-Global-). (accessed on 10 June 2024)
3. A. N. Shiklomanov *et al.*, “Enhancing global change experiments through integration of remote-sensing techniques,” *Frontiers in Ecology and the Environment*, vol. 17, no. 4. Wiley Blackwell, pp. 215–224, May 01, 2019. doi: 10.1002/fee.2031.
4. K. Dewi *et al.*, *Teknik Pengumpulan Data dan Preprocessing Citra Satelit*. Jakarta: Badan Pusat Statistik, 2022.
5. S. R. Putri, A. W. Wijayanto, and S. Pramana, “Multi-source satellite imagery and point of interest data for poverty mapping in East Java, Indonesia: Machine learning and deep learning approaches,” *Remote Sens Appl*, vol. 29, p. 100889, 2023, doi: <https://doi.org/10.1016/j.rsase.2022.100889>.
6. S. Hu, J. Qin, J. Ren, H. Zhao, J. Ren, and H. Hong, “Automatic extraction of water inundation areas using Sentinel-1 data for large plain areas,” *Remote Sens (Basel)*, vol. 12, no. 2, Jan. 2020, doi: 10.3390/rs12020243.
7. S. K. McFEETERS, “The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features,” *Int J Remote Sens*, vol. 17, no. 7, pp. 1425–1432, 1996, doi: 10.1080/01431169608948714.
8. H. Xu, “A Study on Information Extraction of Water Body with the Modified Normalized Difference Water Index (MNDWI),” *National Remote Sensing Bulletin*, 2005, [Online]. Available: <https://api.semanticscholar.org/CorpusID:133201908>
9. Y. Pei, “A Study on Information Extraction of Water System in Semi-arid Regions with the Enhanced Water Index (EWI) and GIS Based Noise Remove Techniques,” *Remote Sensing Information*, 2007, [Online]. Available: <https://api.semanticscholar.org/CorpusID:130618762>
10. T. Zhang, H. Ren, Q. Qin, C. Zhang, and Y. Sun, “Surface Water Extraction from Landsat 8 OLI Imagery Using the LBV Transformation,” *IEEE J Sel Top Appl Earth Obs Remote Sens*, vol. 10, no. 10, pp. 4417–4429, 2017, doi: 10.1109/JSTARS.2017.2719029.
11. Z. Zhang *et al.*, “Automated Surface Water Extraction Combining Sentinel-2 Imagery and OpenStreetMap Using Presence and Background Learning (PBL) Algorithm,” *IEEE J Sel Top Appl Earth Obs Remote Sens*, vol. 12, no. 10, pp. 3784–3798, 2019, doi: 10.1109/JSTARS.2019.2936406.



12. Y. Wang, Z. Li, C. Zeng, G.-S. Xia, and H. Shen, “An Urban Water Extraction Method Combining Deep Learning and Google Earth Engine,” *IEEE J Sel Top Appl Earth Obs Remote Sens*, vol. 13, pp. 769–782, 2020, doi: 10.1109/JSTARS.2020.2971783.
13. M. Drusch *et al.*, “Sentinel-2: ESA’s Optical High-Resolution Mission for GMES Operational Services,” *Remote Sens Environ*, vol. 120, pp. 25–36, 2012, doi: <https://doi.org/10.1016/j.rse.2011.11.026>.
14. X. Yang and L. Chen, “Evaluation of automated urban surface water extraction from Sentinel-2A imagery using different water indices,” *J Appl Remote Sens*, vol. 11, no. 2, p. 26016, 2017, doi: 10.1117/1.JRS.11.026016.
15. Z. Wang, J. Liu, J. Li, and D. D. Zhang, “Multi-Spectral Water Index (MuWI): A Native 10-m Multi-Spectral Water Index for accurate water mapping on sentinel-2,” *Remote Sens (Basel)*, vol. 10, no. 10, Oct. 2018, doi: 10.3390/rs10101643.
16. Bayu Anggoro, “Penyebaran Air Permukaan di Indonesia Tidak Merata.” Available online: <https://mediaindonesia.com/nusantara/283818/penyebaran-air-permukaan-di-indonesia-tidak-merata>. (accessed on 9 June 2024).
17. L. Huang *et al.*, “Analysis of desert locust (*Schistocerca gregaria*) suitability in Yemen: an integrated evaluation based on MaxEnt and space–time cube approaches,” *Int J Digit Earth*, vol. 17, no. 1, 2024, doi: 10.1080/17538947.2024.2346266.
18. J. Li, B. Peng, Y. Wei, and H. Ye, “Accurate extraction of surface water in complex environment based on Google Earth Engine and Sentinel-2,” *PLoS One*, vol. 16, no. 6 June, Jun. 2021, doi: 10.1371/journal.pone.0253209.
19. M. Gašparović and S. K. Singh, “Urban surface water bodies mapping using the automatic k-means based approach and sentinel-2 imagery,” *Geocarto Int*, 2022, doi: 10.1080/10106049.2022.2148757.
20. T. D. Acharya, A. Subedi, and D. H. Lee, “Evaluation of water indices for surface water extraction in a landsat 8 scene of Nepal,” *Sensors (Switzerland)*, vol. 18, no. 8, Aug. 2018, doi: 10.3390/s18082580.
21. M. Gašparović, M. Zrinjski, and M. Gudelj, “Automatic cost-effective method for land cover classification (ALCC),” *Comput Environ Urban Syst*, vol. 76, pp. 1–10, Jul. 2019, doi: 10.1016/j.compenurbsys.2019.03.001.
22. S. Pramana, B. Yuniarto, I. Santoso, R. Nooraeni, and L. H. Suadaa, *Data Mining dengan R, Konsep dan Implementasi*. 2023.
23. J. C.-W. Chan and D. Paelinckx, “Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery,” *Remote Sens Environ*, vol. 112, no. 6, pp. 2999–3011, 2008, doi: <https://doi.org/10.1016/j.rse.2008.02.011>.
24. L. Breiman, “Random Forests,” 2001.
25. T. K. Ho, “The random subspace method for constructing decision forests,” *IEEE Trans Pattern Anal Mach Intell*, vol. 20, no. 8, pp. 832–844, 1998, doi: 10.1109/34.709601.
26. V. F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez, “An assessment of the effectiveness of a random forest classifier for land-cover classification,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 67, pp. 93–104, 2012, doi: <https://doi.org/10.1016/j.isprsjprs.2011.11.002>.
27. Humboldt State Univeristy, “Accuracy Metrics.” Available online: [https://gsp.humboldt.edu/olm/Courses/GSP\\_216/lessons/accuracy/metrics.html](https://gsp.humboldt.edu/olm/Courses/GSP_216/lessons/accuracy/metrics.html). (accessed on 9 June 2024).
28. S. Szabó, Z. Gácsi, and B. Balázs, “Specific features of NDVI, NDWI and MNDWI as reflected in land cover categories,” *Landscape & Environment*, vol. 10, no. 3–4, pp. 194–202, Sep. 2016, doi: 10.21120/le/10/3-4/13.
29. L. Szabó, B. Deák, T. Bíró, G. J. Dyke, and S. Szabó, “NDVI as a proxy for estimating sedimentation and vegetation spread in artificial lakes-monitoring of spatial and temporal changes by using satellite images overarching three decades,” *Remote Sens (Basel)*, vol. 12, no. 9, May 2020, doi: 10.3390/RS12091468.
30. G. J. Steeneveld, S. Koopmans, B. G. Heusinkveld, and N. E. Theeuwes, “Refreshing the role of open water surfaces on mitigating the maximum urban heat island effect,” *Landsc Urban Plan*, vol. 121, pp. 92–96, 2014, doi: <https://doi.org/10.1016/j.landurbplan.2013.09.001>.