

Analisis Spasio-Temporal Multi-Sensor NASA Menggunakan Algoritma K-Means untuk Zonasi Potensi Kesehatan Tanaman di Wilayah Sumatera

Yohanes Gerardus Haga Zai¹, Yusrida Jelianti Sihite², Andika Veriando Saragih^{3*}

^{1,2,3}Matematika dan Ilmu Pengetahuan Alam, Ilmu Komputer, Universitas Negeri Medan, Medan, Indonesia

Email: ^{1*}yohanesgerardushagazai@gmail.com, ²yusridajeliantisihite@gmail.com, ^{3*}andikaandikarr@gmail.com

(*Email Corresponding Author: andikaandikarr@gmail.com)

Received: 16 Maret 2026 | Revision: 16 Maret 2026 | Accepted: 15 Juli 2026

Abstrak

Penelitian ini bertujuan untuk menganalisis zonasi potensi kesehatan tanaman di wilayah Pulau Sumatera menggunakan pendekatan analisis spasio-temporal berbasis data satelit multi-sensor. Data yang digunakan meliputi Enhanced Vegetation Index (EVI) dari produk MOD13Q1, Land Surface Temperature (LST) dari produk MOD11A1, serta data curah hujan dari produk GPM IMERG. Seluruh data satelit diproses melalui tahapan pra-pemrosesan yang mencakup ekstraksi data, mosaicking citra, penyalarsan sistem koordinat, resampling spasial, serta pembersihan data untuk menghilangkan piksel anomali. Data yang telah diproses kemudian dianalisis menggunakan algoritma K-Means Clustering untuk mengelompokkan wilayah berdasarkan kemiripan karakteristik lingkungan. Penentuan jumlah cluster optimal dilakukan menggunakan metode Elbow Method, sedangkan kualitas pengelompokan dievaluasi menggunakan Silhouette Score. Hasil penelitian menunjukkan bahwa jumlah cluster optimal adalah enam zona yang merepresentasikan variasi kondisi biofisik wilayah Sumatera, yaitu dataran rendah dengan vegetasi lebat, lahan terbuka atau terdegradasi, dataran basah dengan curah hujan tinggi, dataran pesisir dengan suhu tinggi, wilayah pegunungan dengan suhu rendah, serta wilayah lereng pegunungan sebagai zona transisi. Analisis terhadap lebih dari dua juta piksel observasi menghasilkan nilai Silhouette Score sebesar 0,2739 yang menunjukkan bahwa struktur cluster telah terbentuk dengan cukup baik. Hasil zonasi ini memberikan gambaran spasial mengenai potensi kesehatan tanaman di Pulau Sumatera dan dapat dimanfaatkan sebagai dasar dalam perencanaan pengelolaan lahan serta pengembangan sistem pertanian yang lebih adaptif terhadap kondisi lingkungan.

Kata Kunci: Penginderaan Jauh, EVI, Suhu Permukaan Daratan, Curah Hujan, K-Means Clustering, Kesehatan tanaman.

Abstract

This study aims to analyze the zoning of potential plant health in the Sumatra region using a spatio-temporal analysis approach based on multi-sensor satellite data. The data used include Enhanced Vegetation Index (EVI) from the MOD13Q1 product, Land Surface Temperature (LST) from the MOD11A1 product, and rainfall data from GPM IMERG. All satellite data were processed through several preprocessing stages including data extraction, image mosaicking, coordinate system alignment, spatial resampling, and data cleaning to remove anomalous pixels. The processed dataset was then analyzed using the K-Means Clustering algorithm to group regions based on similarities in environmental characteristics. The optimal number of clusters was determined using the Elbow Method, while clustering performance was evaluated using the Silhouette Score. The results indicate that six clusters represent the optimal zoning structure, reflecting the bio-physical diversity of Sumatra, including dense lowland vegetation, degraded or open land, high-rainfall wet areas, high-temperature coastal plains, cooler mountainous regions, and transitional mountain slopes. Analysis of more than two million valid observation pixels produced a Silhouette Score of 0.2739, indicating a reasonably well-defined cluster structure. The resulting zoning map provides a spatial representation of potential plant health conditions across Sumatra and can serve as a reference for land management planning and the development of environmentally adaptive agricultural systems.

Keywords: Remote Sensing, EVI, Land Surface Temperature, Rainfall, K-Means Clustering, Plant Health.

1. PENDAHULUAN

Sektor pertanian memiliki peranan penting dalam menjaga ketahanan pangan sekaligus mendukung stabilitas ekonomi di berbagai wilayah. Namun demikian, sektor ini saat ini menghadapi berbagai tantangan yang semakin kompleks akibat dinamika perubahan lingkungan global. Peningkatan suhu rata-rata permukaan bumi, perubahan pola curah hujan, serta meningkatnya frekuensi kejadian iklim ekstrem dapat memengaruhi pertumbuhan tanaman dan produktivitas lahan. Variasi kondisi lingkungan tersebut berpotensi menimbulkan stres pada tanaman, menurunkan hasil produksi, bahkan meningkatkan risiko kegagalan panen pada berbagai komoditas pertanian. Kondisi ini menunjukkan bahwa sektor pertanian memerlukan sistem pemantauan lingkungan yang mampu menyediakan informasi secara cepat, akurat, dan berkelanjutan mengenai kondisi vegetasi serta faktor-faktor lingkungan yang memengaruhi pertumbuhan tanaman. Salah satu pendekatan yang banyak digunakan untuk tujuan tersebut adalah teknologi penginderaan jauh (remote sensing), yaitu metode pengamatan permukaan bumi menggunakan sensor yang dipasang pada satelit atau platform udara tanpa kontak langsung dengan objek yang diamati. Melalui teknologi ini, kondisi vegetasi dapat dipantau secara luas dengan memanfaatkan citra satelit yang tersedia secara global [1].

Pendekatan pemantauan pertanian secara konvensional, seperti survei lapangan dan pengumpulan data manual, memiliki beberapa keterbatasan. Selain membutuhkan waktu yang relatif lama, metode tersebut juga memerlukan biaya operasional yang cukup besar dan sulit diterapkan pada wilayah yang memiliki cakupan geografis luas. Akibatnya,

informasi yang diperoleh sering kali tidak mampu merepresentasikan kondisi pertanian secara menyeluruh. Perkembangan teknologi pengamatan bumi atau Earth Observation (EO) memberikan alternatif yang lebih efektif untuk mengatasi keterbatasan tersebut. Data satelit memungkinkan analisis spasial yang mencakup area luas sekaligus analisis temporal yang dapat menggambarkan perubahan kondisi vegetasi dari waktu ke waktu. Selain itu, meningkatnya ketersediaan data satelit yang bersifat terbuka serta dukungan teknologi komputasi modern membuat proses pengolahan data penginderaan jauh menjadi lebih efisien. Oleh karena itu, teknologi EO semakin banyak dimanfaatkan dalam berbagai penelitian yang berkaitan dengan pemantauan lingkungan dan pertanian berbasis data [2].

Dalam kajian penginderaan jauh yang berkaitan dengan vegetasi, analisis indeks vegetasi merupakan pendekatan yang paling umum digunakan. Indeks vegetasi dihitung dari kombinasi nilai reflektansi spektral yang direkam oleh sensor satelit untuk menggambarkan kondisi fisiologis tanaman, tingkat kehijauan kanopi, serta aktivitas fotosintesis vegetasi. Dua indeks yang sering digunakan dalam penelitian lingkungan adalah Normalized Difference Vegetation Index (NDVI) dan Enhanced Vegetation Index (EVI). Kedua indikator tersebut mampu memberikan gambaran mengenai tingkat kesehatan vegetasi sekaligus perubahan dinamika pertumbuhan tanaman dalam suatu wilayah. Keunggulan utama dari data indeks vegetasi adalah kemampuannya menyediakan informasi spasial yang luas serta resolusi temporal yang memungkinkan analisis perubahan vegetasi dalam periode waktu tertentu. Oleh karena itu, pendekatan analisis spasio-temporal sering digunakan untuk memahami dinamika kondisi vegetasi secara lebih komprehensif [3].

Selain kondisi vegetasi, kesehatan tanaman juga dipengaruhi oleh berbagai faktor lingkungan lain, terutama suhu permukaan daratan dan curah hujan. Suhu yang terlalu tinggi dapat meningkatkan tingkat evaporasi serta memicu stres pada tanaman, sedangkan ketersediaan air dari curah hujan berperan penting dalam mendukung proses pertumbuhan tanaman. Oleh karena itu, analisis kondisi vegetasi akan menjadi lebih representatif apabila dilakukan dengan mengintegrasikan berbagai parameter lingkungan secara bersamaan. Beberapa misi satelit yang dikembangkan oleh NASA menyediakan data lingkungan yang dapat dimanfaatkan untuk tujuan tersebut. Sensor Moderate Resolution Imaging Spectroradiometer (MODIS) menyediakan informasi mengenai indeks vegetasi dan suhu permukaan daratan, sedangkan misi Global Precipitation Measurement (GPM) menyediakan estimasi curah hujan global dengan resolusi temporal yang relatif tinggi. Integrasi berbagai sumber data satelit ini dikenal sebagai pendekatan multi-sensor yang memungkinkan analisis kondisi lingkungan dilakukan secara lebih komprehensif [4].

Sejumlah penelitian sebelumnya telah memanfaatkan data penginderaan jauh untuk menganalisis dinamika vegetasi serta dampak perubahan lingkungan terhadap sektor pertanian. Penelitian oleh Harbowo dan Muliawati [5] misalnya menggunakan data indeks vegetasi untuk menganalisis kondisi kekeringan di wilayah Pulau Nusa Penida. Dalam penelitian tersebut digunakan indikator seperti Enhanced Vegetation Index (EVI) dan Standardized Vegetation Index (SVI) untuk memantau perubahan vegetasi dalam periode waktu tertentu. Selain itu, penelitian tersebut juga menerapkan algoritma K-Means Clustering untuk mengelompokkan wilayah berdasarkan tingkat kerentanan terhadap kekeringan. Penelitian lain menunjukkan bahwa citra satelit multi-temporal dapat digunakan untuk mengidentifikasi perubahan tutupan vegetasi melalui teknik klasifikasi citra dan analisis change detection [6].

Pemanfaatan data indeks vegetasi dari sensor MODIS juga telah digunakan dalam berbagai penelitian untuk memantau dinamika vegetasi dalam jangka panjang. Studi yang dilaporkan dalam [12] menggunakan data Enhanced Vegetation Index dari produk MOD13Q1 untuk menganalisis stres vegetasi di wilayah Jerman melalui pendekatan deret waktu selama lebih dari dua dekade. Sementara itu, penelitian lain [13] memanfaatkan data NDVI dan EVI untuk mengkaji dinamika vegetasi pada ekosistem kering menggunakan metode statistik Mann-Kendall test dan Theil-Sen estimator guna mendeteksi tren perubahan vegetasi dalam jangka panjang. Selain itu, penelitian [14] menunjukkan bahwa data EVI dari produk MOD13Q1 dapat dimanfaatkan sebagai indikator dalam memantau kondisi vegetasi dan mendeteksi potensi kekeringan dengan resolusi spasial sekitar 250 meter serta resolusi temporal 16 hari.

Meskipun berbagai penelitian tersebut menunjukkan potensi besar data penginderaan jauh dalam pemantauan vegetasi, sebagian besar studi masih berfokus pada satu parameter lingkungan secara terpisah. Pendekatan yang mengintegrasikan beberapa variabel lingkungan secara simultan masih relatif terbatas, khususnya dalam konteks pemetaan zonasi potensi kesehatan tanaman pada wilayah yang luas. Selain itu, pemanfaatan metode analisis berbasis clustering untuk mengelompokkan wilayah berdasarkan kombinasi beberapa variabel lingkungan masih memerlukan pengembangan lebih lanjut agar dapat menghasilkan informasi spasial yang lebih komprehensif.

Penelitian [15] menjelaskan bahwa algoritma K-Means merupakan salah satu metode clustering tidak terawasi yang banyak digunakan dalam analisis data berskala besar, termasuk pada data penginderaan jauh. Metode ini memiliki keunggulan dalam hal kesederhanaan komputasi dan efisiensi dalam memproses dataset berdimensi tinggi. Namun demikian, K-Means memiliki keterbatasan karena memerlukan penentuan jumlah kluster sebelum proses pengelompokan dilakukan. Oleh karena itu, berbagai metode evaluasi kluster seperti Silhouette Index, Calinski-Harabasz Index, Davies-Bouldin Index, dan Gap Statistic digunakan untuk membantu menentukan jumlah kluster yang paling optimal.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk melakukan analisis spasio-temporal menggunakan data multi-sensor dari NASA yang meliputi indeks vegetasi, suhu permukaan daratan, serta curah hujan untuk mengidentifikasi zonasi potensi kesehatan tanaman di wilayah Sumatera. Analisis dilakukan dengan menerapkan algoritma K-Means Clustering untuk mengelompokkan wilayah berdasarkan karakteristik lingkungan yang serupa. Hasil penelitian diharapkan dapat menghasilkan peta zonasi yang menggambarkan variasi potensi kesehatan tanaman sehingga dapat memberikan informasi yang bermanfaat dalam mendukung pengelolaan lahan serta pengambilan keputusan di sektor pertanian.

2. METODOLOGI PENELITIAN

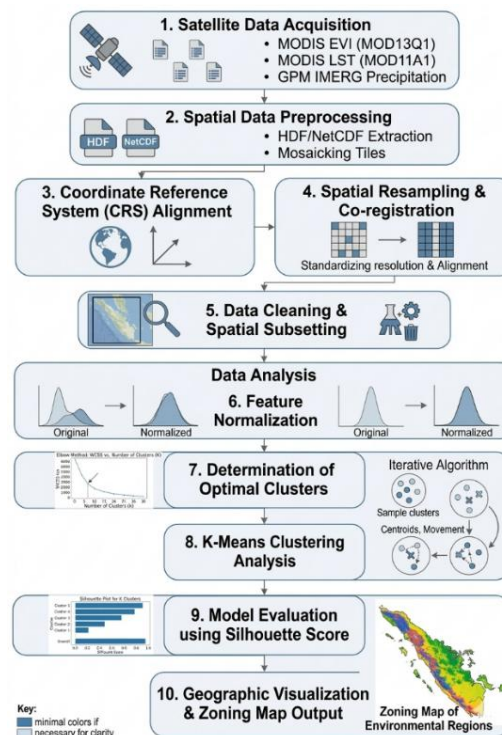
Penelitian ini dilakukan melalui beberapa tahapan utama yang mencakup pengumpulan data satelit, proses pra-pemrosesan data spasial, pembersihan data, pemodelan berbasis *machine learning*, serta visualisasi hasil dalam bentuk peta zonasi. Setiap tahapan dirancang untuk memastikan bahwa data yang digunakan memiliki kualitas yang baik dan dapat dianalisis secara konsisten dalam kerangka analisis spasio-temporal.

2.1 Alur Penelitian

Penelitian ini dilakukan melalui beberapa tahapan yang tersusun secara sistematis untuk menghasilkan analisis zonasi potensi kesehatan tanaman berbasis data satelit. Tahapan penelitian dimulai dari proses akuisisi data satelit, dilanjutkan dengan tahap pra-pemrosesan data spasial, integrasi data multi-sensor, hingga penerapan metode *machine learning* untuk mengelompokkan wilayah berdasarkan karakteristik lingkungan.

Data yang digunakan dalam penelitian ini meliputi indeks vegetasi Enhanced Vegetation Index (EVI) dari produk MOD13Q1, suhu permukaan daratan (Land Surface Temperature/LST) dari produk MOD11A1, serta data curah hujan dari produk GPM IMERG. Seluruh data kemudian melalui proses pra-pemrosesan yang mencakup ekstraksi data, mosaicking citra, penyalarsan sistem koordinat, serta penyesuaian resolusi spasial melalui teknik resampling dan ko-registrasi. Setelah itu dilakukan pembersihan data dan pembatasan wilayah penelitian pada Pulau Sumatera menggunakan metode bounding box.

Selanjutnya, data yang telah diproses dinormalisasi dan dianalisis menggunakan algoritma K-Means Clustering untuk mengelompokkan wilayah berdasarkan kemiripan nilai EVI, LST, dan curah hujan. Penentuan jumlah cluster optimal dilakukan menggunakan metode Elbow Method, sedangkan kualitas hasil clustering dievaluasi menggunakan Silhouette Score. Hasil akhir penelitian divisualisasikan dalam bentuk peta zonasi untuk menggambarkan distribusi spasial potensi kesehatan tanaman di Pulau Sumatera. Diagram alur tahapan penelitian ditunjukkan pada Gambar 1.



Gambar 1. Diagram Alur Penelitian Analisis Zonasi Kesehatan Tanaman Berbasis Data Satelit

2.2 Area Studi dan Periode Penelitian

Penelitian ini dilakukan pada wilayah Pulau Sumatera yang merupakan salah satu kawasan penting dalam sektor pertanian di Indonesia. Wilayah ini memiliki variasi kondisi iklim, topografi, serta karakteristik ekosistem yang beragam sehingga menjadi lokasi yang relevan untuk analisis kondisi vegetasi berbasis data satelit. Analisis dilakukan menggunakan data deret waktu (*time series*) yang mencakup periode dua tahun terakhir. Penggunaan data dengan rentang waktu yang relatif panjang bertujuan untuk memperoleh gambaran kondisi lingkungan yang lebih stabil serta mengurangi pengaruh fluktuasi musiman jangka pendek. Dengan pendekatan tersebut, nilai parameter lingkungan seperti indeks vegetasi, suhu permukaan, serta curah hujan direpresentasikan dalam bentuk nilai rata-rata yang mencerminkan kondisi klimatologi dan biofisik wilayah studi secara lebih representatif. Pendekatan spasio-temporal dalam penelitian ini

memungkinkan identifikasi pola kondisi vegetasi yang tidak hanya bergantung pada distribusi spasial, tetapi juga mempertimbangkan dinamika perubahan yang terjadi sepanjang periode pengamatan.

2.3 Pengumpulan Data

Data yang digunakan dalam penelitian ini berasal dari beberapa produk satelit yang dikembangkan oleh NASA dan mitra internasionalnya. Data tersebut mencakup parameter vegetasi, suhu permukaan daratan, serta curah hujan yang diperoleh dari sensor satelit yang berbeda. Integrasi berbagai sumber data ini bertujuan untuk memperoleh gambaran kondisi lingkungan yang lebih komprehensif.

Tabel 1. Jenis Data Satelit yang Digunakan

Kode Data NASA	Kegunaan Data	Data Spesifik yang Digunakan
MOD13Q1	Mengukur kondisi kesehatan vegetasi melalui indeks vegetasi	Enhanced Vegetation Index (EVI) dari satelit Terra/Aqua MODIS dengan resolusi spasial 250 meter
MOD11A1	Mengukur suhu permukaan daratan	Land Surface Temperature (LST_Day_1km) dengan resolusi spasial 1 km
GPM_3IMERGDF	Mengukur distribusi curah hujan	Data presipitasi dari satelit GPM dengan resolusi spasial sekitar 10 km

Data EVI digunakan sebagai indikator utama untuk menggambarkan kondisi kesehatan vegetasi. Sementara itu, data suhu permukaan daratan digunakan untuk merepresentasikan kondisi termal lingkungan yang dapat memengaruhi pertumbuhan tanaman. Data curah hujan digunakan untuk menggambarkan ketersediaan air pada wilayah studi yang juga menjadi faktor penting dalam pertumbuhan vegetasi.

2.4 Pra-Pemrosesan Data Spasial

Data satelit yang diperoleh dari berbagai sensor memiliki format, resolusi, dan sistem koordinat yang berbeda. Oleh karena itu, diperlukan serangkaian proses pra-pemrosesan untuk memastikan bahwa seluruh data dapat dianalisis secara konsisten dalam satu kerangka spasial yang sama.

- Ekstraksi Format Data Multidimensi:** Data satelit umumnya tidak tersedia dalam bentuk tabel sederhana, melainkan disimpan dalam format multidimensi seperti *Hierarchical Data Format (HDF)* dan *Network Common Data Form (NetCDF)*. Format tersebut menyimpan informasi dalam struktur hierarki yang memuat berbagai variabel ilmiah beserta metadata yang terkait. Pada tahap ini dilakukan proses ekstraksi data untuk mengambil variabel yang relevan dengan penelitian, seperti nilai EVI, suhu permukaan, dan curah hujan.
- Mosaicking Data Satelit:** Wilayah Pulau Sumatera tidak tercakup dalam satu citra satelit tunggal, melainkan tersebar dalam beberapa tile pada sistem grid MODIS. Oleh karena itu, dilakukan proses *mosaicking* untuk menggabungkan beberapa tile citra menjadi satu kesatuan wilayah spasial. Tile yang digunakan antara lain h27v08, h27v09, h28v08, dan h28v09. Proses penggabungan dilakukan dengan manipulasi matriks spasial secara horizontal dan vertikal sehingga menghasilkan satu raster yang mewakili keseluruhan wilayah studi. Pada area laut yang tidak memiliki data vegetasi, nilai *NaN (Not a Number)* disisipkan untuk menjaga konsistensi dimensi matriks.
- Penetapan Sistem Referensi Koordinat:** Data raster yang telah digabungkan kemudian diberikan sistem referensi koordinat (*Coordinate Reference System/CRS*) agar dapat dipetakan secara geografis. Data MODIS dikonfigurasi menggunakan proyeksi Sinusoidal, sedangkan data curah hujan GPM menggunakan sistem koordinat geografis WGS84 (EPSG:4326). Penetapan CRS memungkinkan setiap piksel data memiliki posisi geografis yang jelas pada permukaan bumi.
- Pemilihan Grid Acuan:** Dalam proses integrasi data multi-sensor, data EVI dengan resolusi spasial tertinggi (250 meter) dipilih sebagai grid acuan utama (master grid). Pendekatan ini dilakukan untuk mempertahankan detail spasial yang lebih tinggi dan menghindari kehilangan informasi yang dapat terjadi apabila data beresolusi tinggi diintegrasikan ke resolusi yang lebih kasar.
- 5 Ko-Registrasi dan Penyelarasan Resolusi:** Agar seluruh data memiliki dimensi spasial yang sama, dilakukan proses ko-registrasi dan resampling spasial. Data suhu permukaan dengan resolusi 1 km serta data curah hujan dengan resolusi sekitar 10 km diselaraskan ke resolusi 250 meter mengikuti grid EVI. Selain itu, data curah hujan yang awalnya berada dalam sistem koordinat WGS84 juga mengalami proses reproyeksi agar sesuai dengan sistem proyeksi Sinusoidal yang digunakan oleh data MODIS.
- Transformasi Data ke Format Tabular:** Setelah seluruh data berada dalam dimensi spasial yang seragam, struktur raster kemudian dikonversi menjadi format tabular melalui proses *flattening*. Proses ini mengubah matriks piksel menjadi barisan data satu dimensi sehingga dapat diproses oleh algoritma *machine learning*. Koordinat spasial yang sebelumnya berada dalam sistem proyeksi Sinusoidal kemudian ditransformasikan kembali ke koordinat geografis berupa bujur dan lintang untuk keperluan visualisasi peta akhir.

2.5 Pra-Pemrosesan Data Spasial

Tahap berikutnya adalah melakukan proses pembersihan data untuk menghilangkan piksel yang tidak relevan atau mengandung anomali. Piksel yang memiliki nilai nol pada variabel EVI, suhu permukaan, maupun curah hujan dieliminasi karena berpotensi merepresentasikan area air, awan tebal, atau kesalahan sensor. Selain itu, dilakukan pembatasan wilayah analisis menggunakan metode bounding box agar penelitian hanya berfokus pada wilayah Pulau Sumatera. Batas koordinat yang digunakan adalah bujur 95.0° hingga 106.5° BT dan lintang -6.0° hingga 6.5° LU/LS. Proses ini memastikan bahwa seluruh data yang dianalisis berada dalam cakupan wilayah studi yang telah ditentukan.

2.6 Pemodelan Clustering

Data yang telah dibersihkan kemudian dianalisis menggunakan algoritma *K-Means Clustering* untuk mengelompokkan wilayah berdasarkan kemiripan karakteristik lingkungan. Sebelum proses clustering dilakukan, seluruh variabel dinormalisasi agar memiliki skala yang sebanding. Normalisasi diperlukan karena setiap parameter memiliki satuan yang berbeda, misalnya nilai indeks vegetasi, suhu permukaan, dan curah hujan. Penentuan jumlah cluster tidak dilakukan secara arbitrer, melainkan menggunakan metode *Elbow Method*. Metode ini mengevaluasi perubahan nilai *inertia* atau distorsi pada berbagai jumlah cluster untuk menemukan titik siku pada grafik yang menunjukkan jumlah cluster yang paling optimal.

2.7 Evaluasi Model dan Visualisasi Peta

Untuk menilai kualitas hasil clustering, penelitian ini menggunakan *Silhouette Score* yang mengukur tingkat kesamaan data dalam satu cluster dibandingkan dengan cluster lainnya. Mengingat ukuran dataset yang sangat besar, evaluasi dilakukan menggunakan teknik random sampling agar proses perhitungan tetap efisien. Label cluster yang dihasilkan kemudian dipetakan kembali ke koordinat geografis menggunakan data bujur dan lintang. Proses ini menghasilkan peta zonasi potensi kesehatan tanaman yang menggambarkan distribusi wilayah dengan karakteristik lingkungan yang serupa pada Pulau Sumatera.

3. HASIL DAN PEMBAHASAN

Penelitian ini mengintegrasikan data multi-sensor dari satelit MODIS (EVI dan LST) serta GPM IMERG yang dikumpulkan selama periode pengamatan dua tahun. Batasan wilayah studi dalam penelitian ini ditentukan menggunakan metode *bounding box* yang mencakup koordinat geografis Pulau Sumatera secara menyeluruh. Mengingat letak geografis Pulau Sumatera yang membujur secara diagonal, penggunaan kotak koordinat ini secara logis mencakup sebagian wilayah di luar target utama, seperti Semenanjung Malaya dan sebagian Kepulauan Bangka Belitung. Namun, fokus analisis tetap diprioritaskan pada karakteristik daratan utama.

Seluruh data mentah telah melalui serangkaian tahap pra-pemrosesan, serta pembersihan data. Tahap pembersihan ini krusial untuk mengeliminasi piksel anomali yang disebabkan oleh gangguan atmosfer, tutupan awan tebal, maupun representasi tubuh air. Proses tersebut menghasilkan sebanyak 2.101.456 piksel observasi valid yang memiliki kelengkapan atribut, yang selanjutnya digunakan sebagai basis data input dalam proses zonasi menggunakan algoritma *K-Means Clustering*.

3.1 Karakteristik Variabel Spasial Sumatra

Analisis statistik deskriptif terhadap tiga variabel utama, yaitu Enhanced Vegetation Index (EVI), Land Surface Temperature (LST), dan Global Precipitation Measurement (GPM). Analisis ini bertujuan untuk memberikan gambaran umum mengenai kondisi biofisik wilayah sumatra selama periode pengamatan dua tahun.

Tabel 2. Jenis Data Satelit yang Digunakan

	EVI(indeks)	LST(°C)	GPM(mm/hari)
Min	-0.1373	8.03	5.014583
Mean	0.4884089	28.16048	8.237717
Max	0.769404	40.51000	19.95373
Std	0.07985924	2.617889	1.654598

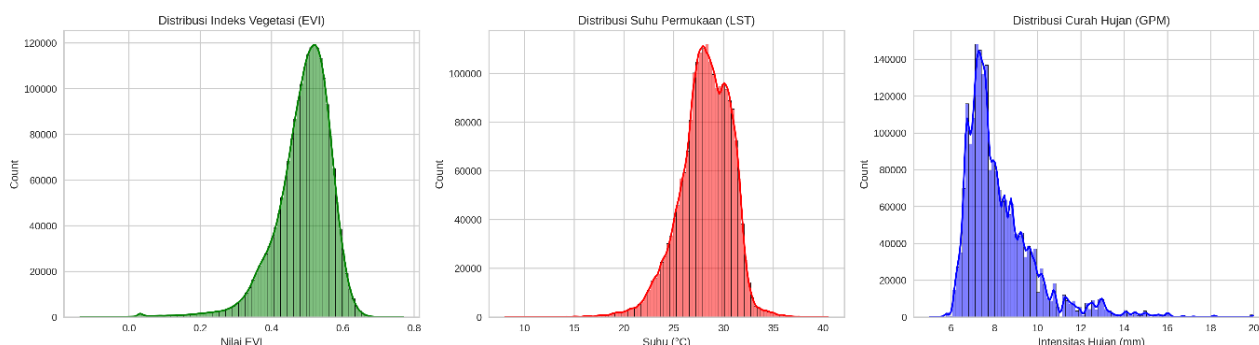
Berdasarkan Tabel 3.1, dapat dianalisis karakteristik masing-masing variabel sebagai berikut:

- Indeks Vegetasi (EVI), nilai rata-rata EVI sebesar 0.4884089 menunjukkan bahwa secara umum wilayah Sumatera memiliki kerapatan vegetasi yang tinggi. Nilai maksimum yang mencapai 0.769404 merepresentasikan area hutan primer atau perkebunan rimbun, sementara nilai minimum -0.1373 umumnya merepresentasikan wilayah dengan tutupan lahan terbuka atau pemukiman.
- Suhu Permukaan Daratan (LST), suhu permukaan rata-rata di wilayah studi tercatat sebesar 28.16048 °C. Rentang suhu yang cukup lebar, antara 8.03°C hingga 40.51000°C, mencerminkan keragaman topografi Pulau Sumatera, di

mana suhu rendah terkonsentrasi di dataran tinggi sepanjang Bukit Barisan dan suhu tinggi mendominasi dataran rendah serta wilayah pesisir.

- c. Curah Hujan (GPM), variabel curah hujan menunjukkan intensitas rata-rata harian sebesar 8.237717 mm/hari. Tingginya standar deviasi pada variabel ini mengindikasikan adanya variasi curah hujan yang signifikan antar wilayah.

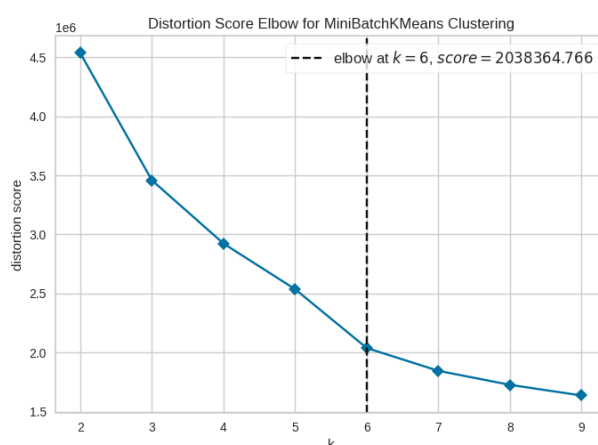
Secara keseluruhan, sebaran nilai pada ketiga variabel ini menunjukkan heterogenitas kondisi alam yang kompleks di Pulau Sumatera.



Gambar 2. Histogram Distribusi Variabel

3.2 Penentuan Jumlah Kluster Optimal (Elbow Method)

Tahapan krusial dalam algoritma *K-Means* adalah menentukan jumlah kluster yang paling optimal untuk merepresentasikan karakteristik data. Pada penelitian ini, penentuan nilai *k* dilakukan menggunakan metode *Elbow*.



Gambar 3. Grafik Elbow

Pada saat mencapai $k = 6$, terjadi penurunan kemiringan yang signifikan (titik siku), di mana penambahan kluster setelah angka tersebut tidak lagi memberikan penurunan distorsi yang berarti. Hal ini menunjukkan titik keseimbangan paling efisien antara kompleksitas model dan kepadatan kluster. Nilai *distortion score* yang berada pada 2.038.364 (dua juta) berkorelasi langsung dengan volume data (*Big Data*) yang diolah, yakni sebanyak 2.133.784 piksel. Skor distorsi merupakan hasil akumulasi jarak kuadrat dari jutaan titik observasi.

3.3 Hasil Zonasi Kesesuaian Lahan

Setelah menetapkan $k = 6$ sebagai jumlah kluster optimal, algoritma *MiniBatchKMeans* mengelompokkan 2,1 juta piksel observasi ke dalam enam zona kesesuaian lahan yang berbeda.

Tabel 3. Profil Karakteristik Biofisik Rata-rata per Zona Cluster

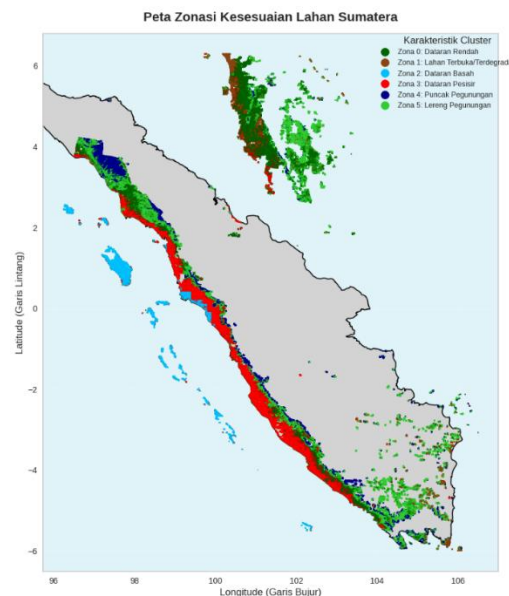
	EVI	LST	GPM
Zona_Cluster			
0	0.555492	29.495106	7.387073
1	0.315791	29.948300	7.751486
2	0.515964	28.668800	12.907007
3	0.508853	30.419240	9.277185
4	0.408552	23.391654	7.734091
5	0.486538	26.835557	7.512328

Berdasarkan Tabel 2, karakteristik spesifik dari setiap zona didefinisikan sebagai berikut:

- Zona 0 (Dataran Rendah), memiliki nilai EVI tertinggi, menunjukkan wilayah dengan tutupan hijau paling lebat.
- Zona 1 (Lahan Terbuka/Terdegradasi): Ditandai dengan EVI terendah dan suhu panas. Zona ini merepresentasikan area dengan vegetasi atau lahan yang mengalami pembukaan.
- Zona 2 (Dataran Basah): Memiliki karakteristik curah hujan paling ekstrem. Wilayah ini adalah area dengan suplai air melimpah yang krusial bagi komoditas yang membutuhkan hidrasi tinggi.
- Zona 3 (Dataran Pesisir): Mencatat suhu permukaan tertinggi. Merupakan zona panas khas dataran rendah pesisir yang memiliki potensi evaporasi tinggi.
- Zona 4 (Puncak Pegunungan): Memiliki suhu terendah. Zona ini merepresentasikan wilayah dataran tinggi di sepanjang rangkaian Bukit Barisan yang memiliki keterbatasan suhu bagi tanaman tropis tertentu.
- Zona 5 (Lereng Pegunungan), memiliki profil suhu sedang dan vegetasi stabil. Zona ini berfungsi sebagai wilayah transisi antara lereng pegunungan dan dataran rendah.

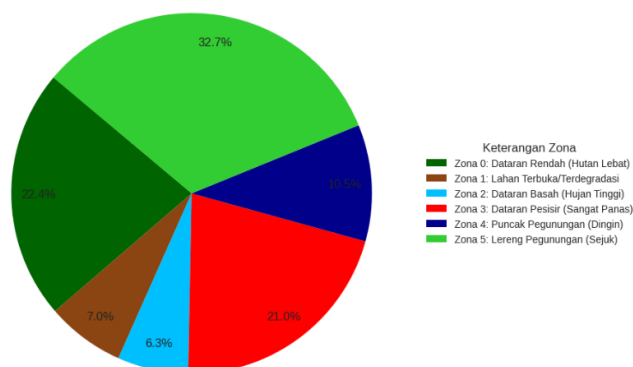
3.4 Visualisasi Spasial dan Distribusi Lahan

Representasi geografis dari hasil *clustering* ditampilkan melalui peta sebaran titik (*scatter map*) dan diagram lingkaran untuk memberikan gambaran komprehensif mengenai distribusi geografis enam zona kesesuaian lahan di Sumatera.



Gambar 4. Gambar peta sebaran titik kesesuaian lahan Sumatera

Distribusi Cluster Zonasi Kesesuaian Lahan Sumatera



Gambar 5. Gambar diagram lingkaran distribusi kesesuaian lahan Sumatera

Setiap zona memiliki pola sebaran yang mengikuti karakteristik biofisik wilayahnya. Zona 0 (Dataran Rendah) mendominasi sebagian besar wilayah dengan tutupan vegetasi paling lebat, disusul oleh Zona 5 (Lereng Pegunungan) sebagai area transisi. Sementara itu, Zona 4 (Puncak Pegunungan) secara konsisten membentuk garis linier yang mengikuti morfologi pegunungan Bukit Barisan. Adapun Zona 2 (Dataran Basah) memperlihatkan konsentrasi di

wilayah-wilayah dengan akumulasi curah hujan tinggi, dan Zona 3 (Dataran Pesisir) mencakup wilayah-wilayah dengan suhu permukaan tertinggi di sepanjang garis pantai.

Munculnya titik data di wilayah luar Sumatera, seperti Semenanjung Malaya dan Bangka Belitung, merupakan konsekuensi dari penggunaan metode *bounding box* dalam ekstraksi koordinat awal. Mengingat posisi Sumatera yang miring secara diagonal, kotak koordinat tersebut secara geometris mencakup area administratif tetangga di sudut-sudutnya. Kedua, adanya area kosong berwarna abu-abu di dalam daratan Sumatera menunjukkan wilayah "No Data". Hal ini disebabkan oleh proses pembersihan data (*data cleaning*) terhadap piksel yang memiliki gangguan atmosfer atau tutupan awan persisten yang tidak dapat ditembus oleh sensor optik satelit selama periode pengamatan. Penghapusan piksel-piksel tidak valid ini dilakukan untuk menjaga akurasi pusat klaster (*centroid*) agar tetap merepresentasikan kondisi lahan yang sebenarnya.

3.5 Evaluasi Performa Model (Silhouette Score)

Untuk memvalidasi kualitas pengelompokan yang dihasilkan oleh algoritma MiniBatchKMeans, dilakukan pengujian menggunakan metrik Silhouette Score. Metrik ini mengukur seberapa mirip sebuah titik data dengan kelompoknya sendiri dibandingkan dengan kelompok lain, dengan rentang nilai antara -1 hingga 1. Mengingat besarnya volume data yang mencapai lebih dari 2,1 juta piksel, perhitungan skor dilakukan menggunakan teknik random sampling terhadap 250.000 titik observasi untuk menjaga efisiensi komputasi.

Berdasarkan pengujian tersebut, model dengan $k=6$ menghasilkan nilai Silhouette Score sebesar 0.2739. Nilai positif menunjukkan bahwa struktur klaster yang terbentuk telah terpartisi dengan cukup baik. Meskipun nilainya tidak mendekati 1, hal ini merupakan fenomena yang wajar dalam analisis data geospasial skala luas. Selain dikarenakan karakteristik variabel alam bersifat kontinu, rendahnya skor tersebut juga dipengaruhi oleh volume data yang sangat besar (*big data*) yang mencapai lebih dari 2 juta observasi, yang secara alami meningkatkan variansi internal di dalam klaster. Faktor lain yang turut berpengaruh adalah kompleksitas topografi Pulau Sumatera yang sangat heterogen, di mana keberadaan mikro-iklim lokal menciptakan titik-titik data yang unik dan saling bersinggungan. Dengan demikian.

3.6 Implikasi untuk Kesehatan Tanaman

Hasil zonasi yang mengintegrasikan variabel EVI, LST, dan GPM memberikan landasan ilmiah untuk memetakan potensi kesehatan tanaman di wilayah Sumatera. Dalam analisis ini, kesehatan tanaman tidak hanya dinilai berdasarkan tingkat kehijauan vegetasi, tetapi juga mempertimbangkan faktor lingkungan yang berperan dalam mendukung atau membatasi pertumbuhan tanaman. Kombinasi antara kondisi vegetasi, suhu permukaan, dan ketersediaan air memberikan gambaran yang lebih komprehensif mengenai kondisi agroekologi suatu wilayah.

Zona 0, yang mencatat nilai rata-rata EVI tertinggi, menunjukkan wilayah dengan potensi kesehatan tanaman yang sangat baik. Nilai EVI yang tinggi mengindikasikan tingkat aktivitas fotosintesis yang optimal serta kepadatan tajuk vegetasi yang relatif tinggi. Kondisi ini biasanya ditemukan pada wilayah yang memiliki kesuburan tanah yang baik, ketersediaan air yang cukup, serta gangguan lingkungan yang relatif rendah. Oleh karena itu, zona ini berpotensi menjadi wilayah yang sangat produktif untuk kegiatan pertanian maupun kehutanan.

Sebaliknya, Zona 1 yang memiliki nilai EVI paling rendah menunjukkan kondisi vegetasi yang relatif jarang atau terdegradasi. Wilayah ini kemungkinan merepresentasikan area dengan aktivitas antropogenik yang tinggi seperti pembukaan lahan, kawasan permukiman, atau wilayah dengan kondisi tanah yang kurang subur. Kondisi vegetasi yang rendah dapat mengindikasikan tingkat kesehatan tanaman yang kurang optimal sehingga memerlukan perhatian lebih dalam pengelolaan lahan dan konservasi vegetasi.

Zona 2 menunjukkan karakteristik wilayah dengan tingkat curah hujan yang paling tinggi dibandingkan zona lainnya. Ketersediaan air yang melimpah dapat memberikan keuntungan bagi tanaman yang membutuhkan suplai air yang tinggi, seperti tanaman padi dan beberapa komoditas perkebunan tropis. Namun demikian, curah hujan yang terlalu tinggi juga berpotensi menimbulkan risiko seperti genangan air, erosi tanah, serta meningkatnya kelembapan yang dapat memicu perkembangan penyakit tanaman.

Pada Zona 3, suhu permukaan daratan tercatat paling tinggi dibandingkan zona lainnya. Kondisi suhu yang tinggi dapat meningkatkan laju evapotranspirasi tanaman sehingga kebutuhan air tanaman menjadi lebih besar. Apabila tidak diimbangi dengan ketersediaan air yang cukup, kondisi ini dapat memicu stres panas pada tanaman yang pada akhirnya berdampak pada penurunan produktivitas tanaman. Oleh karena itu, wilayah dengan karakteristik seperti ini memerlukan strategi pengelolaan air yang lebih baik agar kondisi tanaman tetap stabil.

Zona 4 menunjukkan karakteristik suhu permukaan yang paling rendah dibandingkan zona lainnya. Kondisi ini umumnya ditemukan pada wilayah dataran tinggi atau pegunungan yang memiliki suhu udara relatif lebih sejuk. Lingkungan dengan suhu rendah dapat memberikan kondisi yang ideal bagi beberapa jenis tanaman hortikultura yang memerlukan suhu dingin untuk mendukung proses pertumbuhan dan pembungaan. Namun demikian, suhu yang terlalu rendah juga dapat membatasi pertumbuhan tanaman tropis tertentu yang lebih adaptif terhadap suhu hangat.

Sementara itu, Zona 5 memiliki nilai variabel yang relatif moderat di antara zona lainnya. Karakteristik ini menunjukkan bahwa wilayah tersebut berada pada kondisi lingkungan yang cukup stabil dengan tingkat vegetasi yang baik serta suhu dan curah hujan yang tidak terlalu ekstrem. Zona ini dapat dianggap sebagai wilayah transisi yang memiliki keseimbangan antara faktor vegetasi, suhu, dan ketersediaan air. Kondisi lingkungan yang relatif stabil ini berpotensi mendukung berbagai jenis komoditas tanaman yang memiliki toleransi lingkungan yang cukup luas.

Secara keseluruhan, hasil analisis zonasi yang diperoleh melalui metode clustering memberikan gambaran yang jelas mengenai variasi kondisi lingkungan di Pulau Sumatera. Informasi ini dapat dimanfaatkan sebagai dasar dalam perencanaan pengelolaan lahan, pemantauan kondisi vegetasi, serta pengembangan strategi pertanian yang lebih adaptif terhadap kondisi lingkungan setempat. Pendekatan berbasis data satelit dan machine learning yang digunakan dalam penelitian ini juga menunjukkan potensi yang besar dalam mendukung sistem pemantauan lingkungan secara berkelanjutan pada skala wilayah yang luas.

Dengan demikian, integrasi antara teknologi penginderaan jauh dan metode analisis data modern dapat memberikan kontribusi yang signifikan dalam memahami dinamika lingkungan dan kondisi kesehatan vegetasi di wilayah tropis. Pendekatan ini tidak hanya bermanfaat dalam konteks penelitian akademik, tetapi juga dapat mendukung pengambilan keputusan dalam pengelolaan sumber daya alam dan pembangunan sektor pertanian yang berkelanjutan.

4. KESIMPULAN

Berdasarkan hasil analisis yang telah dilakukan, penelitian ini berhasil mengintegrasikan data satelit multi-sensor untuk mengidentifikasi zonasi potensi kesehatan tanaman di wilayah Pulau Sumatera. Data yang digunakan meliputi indeks vegetasi Enhanced Vegetation Index (EVI), suhu permukaan daratan (Land Surface Temperature/LST), serta curah hujan dari produk satelit NASA. Melalui proses pra-pemrosesan data yang meliputi mosaicking citra, penyelarasan sistem koordinat, resampling spasial, serta pembersihan data, diperoleh lebih dari dua juta piksel observasi valid yang kemudian dianalisis menggunakan algoritma K-Means Clustering. Penentuan jumlah kluster optimal menggunakan metode Elbow Method menunjukkan bahwa enam kluster merupakan konfigurasi yang paling sesuai untuk merepresentasikan variasi karakteristik lingkungan di wilayah Sumatera. Hasil pengelompokan tersebut menghasilkan enam zona utama yang memiliki karakteristik biofisik yang berbeda, yaitu wilayah dataran rendah dengan vegetasi lebat, lahan terbuka atau terdegradasi, wilayah dengan curah hujan tinggi, wilayah pesisir dengan suhu permukaan yang relatif tinggi, wilayah pegunungan dengan suhu yang lebih rendah, serta zona transisi pada lereng pegunungan. Evaluasi model menggunakan Silhouette Score menghasilkan nilai sebesar 0,2739 yang menunjukkan bahwa struktur kluster yang terbentuk memiliki kualitas pengelompokan yang cukup baik, meskipun variasi lingkungan alami dan kompleksitas topografi wilayah menyebabkan adanya tumpang tindih karakteristik antar wilayah. Secara keseluruhan, hasil penelitian ini menunjukkan bahwa integrasi data penginderaan jauh dan metode machine learning mampu memberikan gambaran spasial yang komprehensif mengenai kondisi lingkungan yang memengaruhi kesehatan tanaman. Informasi zonasi yang dihasilkan dapat dimanfaatkan sebagai dasar dalam pemantauan kondisi vegetasi, perencanaan pengelolaan lahan, serta pengembangan strategi pertanian yang lebih adaptif terhadap perubahan lingkungan. Selain itu, pendekatan ini juga menunjukkan potensi besar dalam mendukung sistem pemantauan pertanian berbasis data pada skala wilayah yang luas secara lebih efisien dan berkelanjutan.

REFERENCES

- [1] Vidican, R., Mălinaș, A., Ranta, O., Moldovan, C., Marian, O., Ghețe, A., ... & Cătunescu, G. M. (2023). Using remote sensing vegetation indices for the discrimination and monitoring of agricultural crops: a critical review. *Agronomy*, 13(12), 3040. <https://doi.org/10.3390/agronomy13123040>
- [2] Nakalembe, C., Becker-Reshef, I., Bonifacio, R., Hu, G., Humber, M. L., Justice, C. J., ... & Sanchez, A. (2021). A review of satellite-based global agricultural monitoring systems available for Africa. *Global food security*, 29, 100543. <https://doi.org/10.1016/j.gfs.2021.100543>
- [3] Vidican, R., Mălinaș, A., Ranta, O., Moldovan, C., Marian, O., Ghețe, A., ... & Cătunescu, G. M. (2023). Using remote sensing vegetation indices for the discrimination and monitoring of agricultural crops: a critical review. *Agronomy*, 13(12), 3040. <https://doi.org/10.3390/agronomy13123040>
- [4] Zhan, C., Chen, Y., Yang, K., Zhou, X., Jiang, Y., Ling, X., ... & Yang, H. (2023). First evaluation of GPM-Era satellite precipitation products with new observations on the western Tibetan Plateau. *Atmospheric Research*, 283, 106559. <https://doi.org/10.1016/j.atmosres.2022.106559>
- [5] Harbowo, D. G., & Muliawati, T. (2023, December). A decade drought monitoring through enhanced and standardized vegetation index in isolated karst environment: Nusa Penida Island, Bali, Indonesia. In *IOP Conference Series: Earth and Environmental Science* (Vol. 1287, No. 1, p. 012033). IOP Publishing. 10.1088/1755-1315/1287/1/012033
- [6] Almalki, R., Khaki, M., Saco, P. M., & Rodriguez, J. F. (2022). Monitoring and mapping vegetation cover changes in arid and semi-arid areas using remote sensing technology: A review. *Remote Sensing*, 14(20), 5143. <https://doi.org/10.3390/rs14205143>
- [7] Kartal, S., Iban, M. C., & Sekertekin, A. (2024). Next-level vegetation health index forecasting: A ConvLSTM study using MODIS Time Series. *Environmental Science and Pollution Research*, 31(12), 18932-18948. <https://doi.org/10.1007/s11356-024-32430-x>

- [8] Ma, Y., He, T., McVicar, T. R., Liang, S., Liu, T., Peng, W., ... & Tian, F. (2024). Quantifying how topography impacts vegetation indices at various spatial and temporal scales. *Remote Sensing of Environment*, 312, 114311. <https://doi.org/10.1016/j.rse.2024.114311>
- [9] Benedict, T. D., Brown, J. F., Boyte, S. P., Howard, D. M., Fuchs, B. A., Wardlow, B. D., ... & Evenson, K. A. (2021). Exploring VIIRS continuity with MODIS in an expedited capability for monitoring drought-related vegetation conditions. *Remote Sensing*, 13(6), 1210. <https://doi.org/10.3390/rs13061210>
- [10] Dhillon, M. S., Dahms, T., Kuebert-Flock, C., Rummler, T., Arnault, J., Steffan-Dewenter, I., & Ullmann, T. (2023). Integrating random forest and crop modeling improves the crop yield prediction of winter wheat and oil seed rape. *Frontiers in Remote Sensing*, 3, 1010978. <https://doi.org/10.3389/frsen.2022.1010978>
- [11] Yang, J., Wu, T., Sun, X., Liu, K., Farhan, M., Zhao, X., ... & Wang, S. (2024). Global 24 solar terms phenological MODIS normalized difference vegetation index dataset in 2001–2022. *Geoscience Data Journal*, 11(4), 936-947. <https://doi.org/10.1002/gdj3.268>
- [12] Gessner, U., Reinermann, S., Asam, S., & Kuenzer, C. (2023). Vegetation stress monitor—Assessment of drought and temperature-related effects on vegetation in Germany analyzing MODIS time series over 23 years. *Remote Sensing*, 15(22), 5428. <https://doi.org/10.3390/rs15225428>
- [13] Bagheri, R., Ranjbar-Fordoei, A., Mousavi, S. H., & Tahmasebi, P. (2021). Assessment of MODIS-Derived NDVI and EVI for Different Vegetation Types in Arid Region: A Study in Sirjan Plain Catchment of Kerman province, Iran. *Journal of Rangeland Science*, 11(1).
- [14] Tesfaye, A. A., Osgood, D. E., & Aweke, B. G. (2025). Application of a novel vegetation condition index using MODIS EVI for structuring crop index insurance under a smallholder system. *Environmental and Sustainability Indicators*, 26, 100696. <https://doi.org/10.1016/j.indic.2025.100696>
- [15] Ali, I., Rehman, A. U., Khan, D. M., Khan, Z., Shafiq, M., & Choi, J. G. (2022). Model selection using K-means clustering algorithm for the symmetrical segmentation of remote sensing datasets. *Symmetry*, 14(6), 1149. <https://doi.org/10.3390/sym14061149>