

Perbandingan Metode Partisi Machine Learning untuk Kerentanan Kebakaran Lahan Gambut Kalimantan pada Variasi ENSO

Intan Bryliana Putri^{1,*}, Latifah Nurul Qomariyatuzzamzami¹, Purwanti Lelly Sabrina¹, Yahya Darmawan²

¹Program Studi Meteorologi, Sekolah Tinggi Meteorologi Klimatologi dan Geofisika, Tangerang, Indonesia

²Program Studi Klimatologi, Sekolah Tinggi Meteorologi Klimatologi dan Geofisika, Tangerang, Indonesia

Email: ^{1,*}intnbryliana@gmail.com, ¹latifah.zamzami@stmkg.ac.id, ¹purwantilelly.work@gmail.com,

²yahya.darmawan@bmgk.go.id

Email Penulis Korespondensi: intnbryliana@gmail.com*

Submitted: 23/07/2026; Accepted: 15/09/2026; Published: 30/09/2026

Abstrak– Lahan gambut Kalimantan merupakan wilayah rawan kebakaran yang polanya dipengaruhi fase *El Niño -Southern Oscillation* (ENSO). Penelitian ini membandingkan metode partisi acak KFOLD k=5 dan metode partisi spasial BANDS dalam pemodelan kerentanan kebakaran lahan gambut Kalimantan periode 2015 – 2024, menggunakan algoritma *machine learning* tunggal melalui paket ENMTML pada tiga fase ENSO. Random Forest dengan KFOLD mencapai AUC 0,999 pada El Niño, namun diagnostik kNNNDM menunjukkan statistik Wasserstein KFOLD selalu lebih tinggi dibanding BANDS yang berkisar 90.218-121.109 meter berbanding hanya 19.096-77.225 meter, sehingga performa tinggi KFOLD kurang merepresentasikan kondisi prediksi sesungguhnya. Pada metode BANDS, algoritma terbaik berbeda tiap fase yaitu *Maximum Entropy default* dengan AUC 0,929 pada El Niño, *Boosted Regression Trees* dengan AUC 0,983 pada Netral, dan *Random Forest* dengan AUC 0,870 pada La Niña. Kalimantan Tengah dan Kalimantan Selatan menunjukkan proporsi kerentanan tinggi hingga sangat tinggi terbesar dibanding wilayah lain pada fase ENSO yang berbeda. Variasi ENSO tidak berperan signifikan pada kerentanan kebakaran lahan gambut di Kalimantan dibandingkan dengan aktivitas konveksi lokal dan prediktor kelembaban lingkungan. Pemilihan skema partisi yang memperhitungkan struktur spasial terbukti penting untuk menghasilkan peta kerentanan yang lebih realistis dan berpotensi mendukung upaya konservasi lahan gambut yang lebih sesuai dengan karakteristik lingkungan setempat.

Kata Kunci: *machine learning*; kNNNDM; kerentanan kebakaran; ENSO; gambut tropis

Abstract– Kalimantan's peatlands are fire-prone areas whose patterns are influenced by the El Niño-Southern Oscillation (ENSO) phases. This study compared the random K-fold (k=5) partitioning method and the spatial BANDS partitioning method in modeling peatland fire vulnerability in Kalimantan for the 2015–2024 period, using a single machine learning algorithm through the ENMTML package across three ENSO phases. Random Forest with K-fold achieved an AUC of 0.999 during El Niño, yet kNNNDM diagnostics showed that the Wasserstein statistic for K-fold was consistently higher than that of BANDS, ranging from 90,218 to 121,109 meters compared to only 19,096 to 77,225 meters, indicating that K-fold's high performance did not accurately reflect actual predictive conditions. Under the BANDS method, the best-performing algorithm differed across phases, with default Maximum Entropy achieving an AUC of 0.929 during El Niño, Boosted Regression Trees achieving an AUC of 0.983 during Neutral conditions, and Random Forest achieving an AUC of 0.870 during La Niña. Central Kalimantan and South Kalimantan showed the largest proportions of high to very high vulnerability compared to other regions across the different ENSO phases. ENSO variation did not play a significant role in peatland fire vulnerability in Kalimantan compared to local convective activity and environmental humidity predictors. Selecting a partitioning scheme that accounts for spatial structure proved essential for producing more realistic vulnerability maps, with potential to support peatland conservation efforts better suited to local environmental characteristics.

Keywords: machine learning; kNNNDM; fire vulnerability; ENSO; tropical peatland

1. PENDAHULUAN

Lahan gambut menyimpan cadangan karbon dalam jumlah besar yang dapat terlepas ke atmosfer saat terbakar [1]. Penurunan muka air tanah akibat drainase meningkatkan paparan gambut terhadap oksigen, sehingga ketika terbakar akan melepaskan karbon yang sudah tersimpan lama dan membuat lahan gambut tropis menjadi sumber emisi yang signifikan [2], [3]. Kebakaran hutan dan lahan pada tahun 2015 membakar sekitar 2,61 juta hektare di Indonesia, dengan sekitar 618.574 hektare di antaranya merupakan lahan gambut. Area terbakar terluas tercatat di Sumatera Selatan, Kalimantan Tengah, Kalimantan Timur, Kalimantan Selatan, dan Papua [4].

Skala, frekuensi, dan intensitas kebakaran di Kalimantan tidak lepas dari pengaruh *El Niño-Southern Oscillation* (ENSO) [5]. ENSO terdiri atas tiga fase, yaitu El Niño, Netral, dan La Niña yang dibedakan berdasarkan anomali pemanasan suhu permukaan laut di Pasifik ekuatorial [6]. Fase El Niño cenderung menunda musim hujan serta memperpanjang musim kemarau yang meningkatkan risiko kebakaran, sementara fase La Niña cenderung memajukan musim hujan serta menurunkan aktivitas kebakaran [7]. Pola kejadian kebakaran berasosiasi dengan karakteristik lingkungan tertentu, sehingga memerlukan tersedianya peta kerentanan kebakaran lahan gambut yang dapat menjadi dasar bagi upaya mitigasi dan konservasi yang lebih terarah [8], [9]. Kebutuhan tersebut dapat dijawab melalui pendekatan yang menghubungkan titik kejadian kebakaran dengan sejumlah parameter karakteristik lingkungan di sekitarnya.

Kerangka konseptual *Ecological Niche Model* (ENM) memodelkan hubungan antara kemunculan suatu spesies dengan variabel lingkungan yang memengaruhinya [10], [11]. Prinsip korelatif yang sama telah diadaptasi ke dalam pemodelan kerentanan bencana seperti tanah longsor dan kebakaran hutan dengan menggunakan berbagai pendekatan *machine learning* [11], [12]. Pemilihan algoritma *machine learning* yang digunakan memerlukan evaluasi secara spesifik berdasarkan data dan karakteristik wilayah penelitian [13]. Selain itu, pendekatan *ensemble* yang menggabungkan beberapa algoritma sekaligus berpotensi memberikan kinerja superior dalam hal akurasi dan kemampuan generalisasi dengan efektivitas yang bergantung pada pemilihan metode dan kualitas data dengan evaluasi lanjutan [14]. Perbandingan metode pemodelan *Generalised Additive Models* (GAM) dengan metode *machine learning* yang telah dikalibrasi untuk memprediksi kejadian kebakaran hutan harian di Kanada menunjukkan GAM memiliki kinerja yang setara dengan model *machine learning* namun lebih unggul dalam hal interpretabilitas [15]. Penerapan algoritma *Generalized Linear Model* (GLM) dalam empat metode *ensemble* untuk memprediksi kerentanan bahaya kebakaran di Iran menunjukkan peningkatan performa dibandingkan algoritma tunggal tanpa *ensemble* GLM [16]. Penelitian yang dilakukan di Sumatra memodelkan kerentanan lahan gambut terhadap kebakaran pada tiga fase ENSO dengan menggabungkan algoritma *random forest*, *support vector machine*, *maximum entropy*, dan *boosted regression tree* melalui paket ENMTML menemukan bahwa model *ensemble* memberikan performa yang lebih baik dibanding algoritma tunggal [17]. Meskipun demikian, paket ENMTML yang digunakan oleh penelitian tersebut memiliki sejumlah parameter yang masih memerlukan penyesuaian dan pengujian lebih lanjut, salah satunya adalah metode partisi data [18].

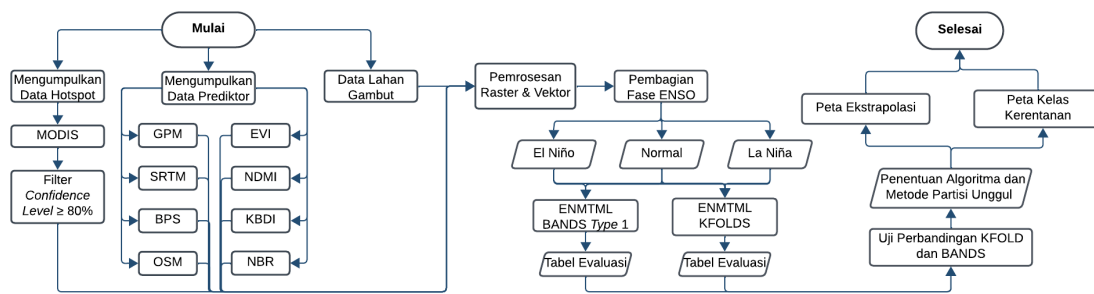
Pemilihan metode partisi menyangkut validitas estimasi performa model. Bahn dan McGill menguji empat skema partisi pada 79 spesies burung di Amerika Utara, dan menemukan bahwa tingkat independensi data uji dan data latih yang lebih tinggi menghasilkan estimasi akurasi yang lebih rendah namun lebih realistis, karena partisi acak sederhana menghasilkan estimasi inflasi akibat autokorelasi spasial [19]. Pengembang paket *blockCV* menegaskan hal serupa dalam konteks pemodelan distribusi spesies, dengan menyatakan bahwa partisi acak konvensional dapat menyebabkan estimasi galat yang lebih rendah dari sebenarnya dan berujung pada pemilihan model yang keliru [20]. Percobaan dengan menggunakan data penginderaan jauh juga menunjukkan bahwa partisi non-spasial cenderung melebih-lebihkan kemampuan generalisasi model, sementara *spatial leave-one-out cross-validation* memberikan estimasi galat yang tidak bias [21]. Percobaan yang membandingkan sepuluh metode pemodelan distribusi spesies menemukan bahwa metode yang fleksibel tetap tampil baik ketika diuji pada data yang dipisah secara spasial, sehingga tidak menunjukkan penurunan performa setajam yang diperkirakan oleh studi-studi sebelumnya [22]. Perbedaan ini diperjelas oleh penelitian yang menguji berbagai ukuran dan jumlah blok pada pemodelan lingkungan laut dan menemukan bahwa ukuran blok merupakan faktor metodologis paling menentukan, di mana partisi spasial dengan blok yang terlalu besar terkadang dapat melebih-lebihkan galat alih-alih memberikan estimasi yang lebih konservatif [23]. Perbedaan temuan antara kedua kelompok penelitian ini menunjukkan bahwa perbedaan performa yang muncul antara partisi acak dan partisi spasial pada suatu model tidak dapat langsung disimpulkan sebagai bukti realisme partisi spasial semata, tanpa mempertimbangkan pula karakteristik data, jumlah kelompok, dan ukuran blok yang digunakan dalam partisi tersebut. Untuk menjawab persoalan ini, dikembangkan diagnostik *k-fold Nearest Neighbour Distance Matching* (kNNDM) [24]. Metode ini merupakan pengembangan dari pendekatan *Nearest Neighbour Distance Matching Leave-One-Out Cross-Validation* (NNDM LOO CV) yang sebelumnya diusulkan untuk validasi peta pada data kontinu [25]. kNNDM mulai diadopsi sebagai kategori evaluasi partisi yang adaptif terhadap domain prediksi (*prediction-domain adaptive evaluation*) dalam pemodelan lingkungan berbasis spasial [26].

Berdasarkan uraian di atas, penelitian pemodelan kerentanan lahan gambut terhadap kebakaran pada berbagai fase ENSO yang dilakukan di Sumatera dan Riau menggunakan paket ENMTML menyediakan opsi partisi acak maupun partisi spasial, namun pemilihan metode partisi ini belum dijelaskan maupun diuji secara eksplisit [18]. Meskipun demikian, diagnostik kNNDM yang dapat menjawab persoalan pemilihan metode partisi tersebut secara empiris belum pernah diterapkan pada penelitian pemodelan kerentanan di Indonesia, termasuk pada kasus kerentanan kebakaran lahan gambut. Oleh karena itu, penelitian ini bertujuan membandingkan performa algoritma tunggal dan bukan *ensemble* dengan pemodelan *niche* ekologi dalam memetakan kerentanan kebakaran lahan gambut di Kalimantan pada tiga fase ENSO menggunakan metode partisi data acak (KFOLD) dan spasial (BANDS) dalam kerangka ENMTML dan menilai kesesuaian metode partisi terhadap kondisi prediksi sesungguhnya melalui diagnostik kNNDM. Hasil penelitian ini diharapkan dapat mengurangi risiko ekstrapolasi yang membiaskan hasil pelatihan model, sehingga peta kerentanan kebakaran lahan gambut yang dihasilkan dapat mendukung upaya konservasi lahan gambut yang lebih sesuai dengan karakteristik lingkungan setempat.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini berlokasi di kawasan lahan gambut Kalimantan yang tersebar di lima provinsi yaitu Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, dan Kalimantan Utara. Proses yang dilakukan memiliki tahapan yang saling mempengaruhi hingga hasil akhir penelitian. Alur tahapan ditunjukkan melalui Gambar 1.



Gambar 1. Diagram Alur Penelitian Perbandingan Metode Partisi ENMTML

2.2 Data dan Sumber Data

Data yang digunakan berada dalam periode 2015 – 2024 dan dikelompokkan menjadi dua kategori yaitu data titik kejadian kebakaran dan data prediktor lingkungan. Data prediktor lingkungan terbagi ke dalam faktor hidrometeorologi yaitu presipitasi dan KBDI, faktor topografi yang memuat elevasi dan kemiringan lahan, faktor antropogenik untuk meninjau keterkaitan aktivitas manusia melalui jarak ke jalan, jarak ke hutan, dan jarak ke permukiman, dan faktor biofisik vegetasi dengan memanfaatkan citra satelit untuk indeks EVI, NDMI, dan NBR. Rincian sumber setiap jenis data disajikan pada Tabel 1.

Tabel 1. Data Masukan Model

Jenis Data	Sumber	Produk	Format	Resolusi
<i>Hotspot</i> (titik kejadian kebakaran)	FIRMS (NASA)	<i>MODIS Collection 6.1 (Terra dan Aqua)</i>	.csv	500 m
Presipitasi	GPM (NASA)	<i>IMERG Final Precipitation L3 1 day</i>	NetCDF4	0,1° (~11 km)
Sebaran lahan gambut	<i>Global Peatland Database (GEE)</i>	Lahan Gambut	GeoTIFF	1 km
Elevasi dan Kemiringan Lahan	SRTM v3 (NASA/USGS/JPL-Caltech)	DEM	GeoTIFF	30 m
Jarak ke jalan, jarak ke hutan, jarak ke permukiman	<i>OpenStreetMap (GEOFABRIK)</i>	<i>roads, forest, places</i>	Shapefile	-
Kepadatan populasi	BPS	Kepadatan Populasi Provinsi	.csv	-
<i>Enhanced Vegetation Index (EVI)</i>	<i>MOD09GA.061 (GEE)</i>	<i>MODIS Surface Reflectance</i>	GeoTIFF	500 m
<i>Normalized Difference Moisture Index (NDMI)</i>	<i>MOD09GA.061 (GEE)</i>	<i>MODIS Surface Reflectance</i>	GeoTIFF	500 m
<i>Normalized Burn Ratio (NBR)</i>	<i>MOD09GA.061 (GEE)</i>	<i>MODIS Surface Reflectance</i>	GeoTIFF	500 m
<i>Keetch-Byram Drought Index (KBDI)</i>	<i>Institute of Industrial Science University Tokyo (GEE)</i>	KBDI	GeoTIFF	4 km

Data BPS yang belum berformat *raster* dikonversi terlebih dahulu ke dalam format tersebut. Data vektor jarak ke jalan, permukiman, dan hutan terlebih dahulu dikonversi menjadi raster biner sebagai lapisan target. Selanjutnya, jarak Euclidean ke fitur terdekat dihitung pada grid terproyeksi sebagai proksi aksesibilitas manusia, sebagaimana lazim digunakan pada pemodelan kerentanan kebakaran lahan gambut [27]. Setelah seluruh data berformat *raster*, resolusi spasialnya diseragamkan menjadi 1 kilometer.

Data *hotspot* kebakaran kemudian disaring berdasarkan tingkat kepercayaan (*confidence level*) dengan hanya mempertahankan titik yang memiliki nilai kepercayaan $\geq 80\%$ [28]. Penyaringan ini bertujuan mengurangi kesalahan deteksi *hotspot* yang bukan berasal dari kebakaran aktual. Titik *hotspot* yang telah disaring selanjutnya dipotong sesuai batas sebaran lahan gambut Kalimantan sehingga hanya titik yang berada di dalam area gambut yang digunakan dalam pemodelan.

2.3 Klasifikasi Fase ENSO

Fase ENSO pada setiap bulan periode penelitian diklasifikasikan menggunakan *Oceanic Niño Index* (ONI) yang diterbitkan oleh *National Oceanic and Atmospheric Administration Climate Prediction Center* (NOAA CPC). ONI

dihitung sebagai rata-rata bergerak tiga bulan dari anomali suhu permukaan laut pada wilayah Niño 3.4. Fase El Niño ditetapkan ketika nilai ONI lebih besar atau sama dengan $+0,5^{\circ}\text{C}$ selama minimal lima musim tumpang tindih berurutan. Fase La Niña ditetapkan ketika nilai ONI lebih kecil atau sama dengan $-0,5^{\circ}\text{C}$ dengan durasi yang sama. Fase Netral ditetapkan untuk kondisi di antara kedua ambang tersebut [8]. Penelitian ini mengklasifikasikan fase ENSO sesuai dengan nilai ONI dalam 10 tahun pada periode 2015 – 2024 yang disajikan pada Tabel 2.

Tabel 2. Ringkasan Klasifikasi Fase ENSO Periode 2015 – 2024

Fase	Jumlah Bulan	Jumlah Titik Hotspot
El Niño	45	20.037
La Niña	31	376
Netral	44	10.644

2.4 Pemrosesan Sebelum Pemodelan

Variabel prediktor yang digunakan dalam penelitian ini terdiri atas variabel statis dan variabel dinamis. Variabel statis meliputi elevasi, kemiringan lereng, kepadatan penduduk, jarak ke jalan, jarak ke permukiman, dan jarak ke hutan yang nilainya dianggap tidak berubah signifikan sepanjang periode penelitian. Variabel dinamis meliputi presipitasi, KBDI, EVI, NDMI, dan NBR yang diagregasi menjadi satu nilai representatif untuk setiap fase ENSO pada setiap sel raster, menggunakan rata-rata dari seluruh bulan pada fase yang bersangkutan sesuai Tabel 2 [29]. Pendekatan ini pada dasarnya mengikuti logika umum peta kesesuaian atau kerentanan yang mengasumsikan kondisi historis dapat merepresentasikan kondisi tipikal suatu kategori tanpa mengintegrasikan variabilitas temporal secara eksplisit ke dalam prediksinya [30]. Konsekuensinya, pendekatan ini memiliki keterbatasan berupa hilangnya variasi temporal tahunan. Intensitas suatu fase ENSO tidak seragam dari satu kejadian ke kejadian lain bahkan dalam kategori fase yang sama, sebagaimana ditunjukkan oleh dekomposisi variansi ENSO ke dalam beberapa kategori intensitas yang menemukan perbedaan signifikan antar kejadian dalam kategori fase yang sama pada skala dekadal [31]. Model yang dihasilkan karena itu merepresentasikan kondisi rata-rata umum suatu fase dan bukan kondisi bulan atau tahun spesifik.

2.5 Pemodelan dengan ENMTML

Pemodelan kerentanan kebakaran lahan gambut dilakukan menggunakan paket ENMTML pada perangkat lunak R [18]. Algoritma digunakan dalam penelitian ini yaitu *Maximum Entropy default* (MXD), *Maximum Entropy simple* (MXS), *Support Vector Machine* (SVM), *Random Forest* (RDF), *Generalized Additive Model* (GAM), *Generalized Linear Model* (GLM), dan *Boosted Regression Tree* (BRT) yang telah digunakan pada beberapa penelitian sebelumnya [16], [17], [32].

Sebelum pemodelan, titik *hotspot* pada setiap fase ENSO dijarangkan secara spasial menggunakan metode *cellsize* dalam paket ENMTML dengan jarak penjarangan yang ditentukan otomatis sebesar dua kali ukuran sel prediktor [33]. Pemilihan metode tersebut karena pengaruh jarak penjarangan terhadap performa model terbukti tidak seragam antar kasus dan justru dapat menurunkan performa model jika jarak dipilih terlalu besar tanpa dasar yang terkait langsung dengan struktur data [34]. Penjarangan spasial maupun lingkungan cenderung menurunkan performa model dibanding data tanpa penjarangan, dengan penurunan yang meningkat seiring membesarnya jarak penjarangan [35]. Tabel 3 merangkum jumlah sebaran *hotspot* sebelum dan setelah dilakukan penjarangan spasial.

Tabel 3. Jumlah Sebaran *Hotspot* Sebelum dan Sesudah Penjarangan Spasial

Fase	Sebelum	Sesudah
El Niño	20.037	11.465
La Niña	376	314
Netral	10.644	5.419

ENMTML dijalankan dengan konfigurasi yang ada pada Tabel 4. Titik *pseudo-absences* dialokasikan menggunakan metode terintegrasi yang menggabungkan batasan ruang geografis, kondisi lingkungan, dan pengelompokan *k-means* untuk distribusi titik yang tersebar secara merata dan mewakili variasi kondisi lingkungan [36]. Rasio disesuaikan dengan perbandingan seimbang yaitu 1:1 [37]. Eliminasi variabel prediktor yang berkorelasi dilakukan dengan metode Pearson pada ambang batas 0,7 [38]. Parameter algoritma yang digunakan adalah parameter standar (*default*) dari paket algoritma di R dalam ENMTML [18]. *Mobility-Oriented Parity* (MOP) digunakan untuk analisis ekstrapolasi spasial model [39]. *Random seed* yang dipilih berfungsi mengunci titik awal generator angka acak sehingga proses yang mengandung elemen stokastik (pembagian data, inialisasi model, bootstrap) menghasilkan output yang sama persis setiap kali dijalankan ulang [40].

Model diproses menggunakan dua metode partisi data secara terpisah untuk membandingkan performanya, yaitu *k-fold cross-validation* dengan lima lipatan (KFOLD, $k=5$) dan partisi spasial (BANDS) tipe 1. Partisi KFOLD membagi data secara acak menjadi lima subset dengan proporsi yang seimbang [41]. Partisi BANDS membagi data berdasarkan struktur geografis wilayah kalibrasi. Pembagian blok geografis ini dioptimalkan menggunakan Indeks Moran untuk mengukur autokorelasi spasial antarblok. Kriteria *Multivariate Environmental Similarity Surface* (MESS) turut digunakan untuk menilai kemiripan kondisi lingkungan antarblok yang terbentuk.

Kedua kriteria tersebut dioptimalkan bersama dengan batasan keseimbangan jumlah sampel antartpartisi, sehingga blok yang dihasilkan tidak didominasi oleh satu wilayah kecil saja [15], [16]. Berbeda dengan KFOLD yang jumlah lipatnya ($k=5$) memang ditentukan langsung oleh pengguna lewat argumen *folds*, metode BANDS pada ENMTML tidak memiliki argumen serupa untuk mengatur jumlah subset. Perbedaan struktur ini bisa berpengaruh pada estimasi performa. Dengan hanya satu subset uji, hasil evaluasi pada BANDS lebih bergantung pada komposisi data di subset tunggal tersebut. Sementara pada KFOLD, hasil evaluasi merupakan rata-rata dari lima subset uji yang berbeda, sehingga variasi acak dari satu subset saja tidak terlalu memengaruhi hasil akhir [41].

Evaluasi performa model menggunakan metrik *Area Under the Curve* (AUC) dan *True Skill Statistic* (TSS) karena mewakili dimensi evaluasi berbeda yang saling melengkapi [42]. *Area Under the ROC Curve* (AUC) mengukur kemampuan model dalam membedakan area dengan dan tanpa kejadian kebakaran di seluruh rentang ambang batas klasifikasi. Nilai AUC berkisar dari 0 hingga 1, di mana nilai mendekati 1 mengindikasikan performa yang sangat baik [43]. *True Skill Statistic* (TSS) merupakan metrik yang menggabungkan *sensitivity* dan *specificity* dengan cara yang seimbang dan tidak terpengaruh oleh prevalensi kejadian dalam dataset [44]. TSS dihitung dengan

$$TSS = Sensitivity + Specificity - 1 \quad (1)$$

Nilai TSS berkisar dari -1 hingga +1 dimana $TSS = 1$ mengindikasikan prediksi sempurna, $TSS = 0$ mengindikasikan prediksi tidak lebih baik dari tebakan acak, dan $TSS < 0$ mengindikasikan prediksi lebih buruk dari tebakan acak. Setiap kombinasi algoritma dan metode partisi dijalankan sebanyak tiga kali replikasi untuk masing-masing fase ENSO. Nilai AUC dan TSS yang dilaporkan merupakan nilai rata-rata dari ketiga replikasi tersebut. Perhitungan rata-rata ini dilakukan secara otomatis oleh ENMTML.

Tabel 4. Konfigurasi ENMTML

Parameter	Metode
<i>Pseudo-absence</i>	Kombinasi geografis, lingkungan, dan pengelompokan <i>k-means</i>
Rasio <i>presence : absence</i>	1:1
Kolinearitas prediktor	Pearson dengan ambang batas $> 0,7$
Penjarangan spasial	<i>cellsize</i>
Metode partisi	KFOLD <i>fold</i> 5 BANDS <i>type</i> 1
Algoritma	<i>Maximum Entropy default</i> (MXD) <i>Maximum Entropy simple</i> (MXS) <i>Support Vector Machine</i> (SVM) <i>Random Forest</i> (RDF) <i>Generalized Additive Model</i> (GAM) <i>Generalized Linear Model</i> (GLM) <i>Boosted Regression Tree</i> (BRT)
Metrik evaluasi	AUC dan TSS
Ekstrapolasi	<i>Mobility-Oriented Parity</i> (MOP)
<i>Random seed</i>	<i>set.seed</i> (42)

2.6 Perbandingan Metode Partisi Acak dan Spasial

Perbandingan KFOLD dan BANDS dilakukan menggunakan diagnostik *k-fold Nearest Neighbour Distance Matching* (kNNDM). Metode ini merupakan pengembangan dari *Nearest Neighbour Distance Matching Leave-One-Out Cross-Validation* (NNDM LOO CV) yang mencocokkan distribusi jarak *nearest-neighbour* selama validasi silang dengan distribusi jarak pada kondisi prediksi sesungguhnya, namun menggunakan algoritma *greedy* yang mengeluarkan observasi satu per satu sehingga tidak efisien untuk data berskala besar [25]. Konsep tersebut lalu diadaptasi ke dalam skema *k-fold* menggunakan algoritma *clustering* (*hierarchical* atau *k-means*) untuk mengelompokkan titik latih menjadi sejumlah *cluster* yang kemudian digabungkan menjadi *k* lipatan, sehingga metode ini dapat diterapkan pada dataset besar dengan biaya komputasi yang jauh lebih rendah dibanding versi *leave-one-out* [24].

Kesesuaian antara kondisi validasi silang dan kondisi prediksi dinilai menggunakan statistik *Wasserstein* (W) dengan

$$W = \int |\hat{G}_j^*(r, L) - \hat{G}_{ij}(r)| dr \quad (2)$$

dengan $\hat{G}_{ij}(r)$ adalah fungsi distribusi kumulatif empiris (*empirical cumulative distribution function*/ECDF) jarak *nearest-neighbour* antara titik prediksi dan titik latih pada domain area prediksi lahan gambut Kalimantan, dirumuskan sebagai

$$\hat{G}_{ij}(r) = \left(\frac{1}{n_i}\right) \sum_i 1\{d_{ij} \leq r\} \quad (3)$$

dengan d_{ij} adalah jarak dari titik prediksi ke titik latih terdekat. $\hat{G}_j^*(r, L)$ adalah ECDF jarak *nearest-neighbour* antara data uji dan data latih untuk skema partisi *L* yang diuji, dirumuskan sebagai

$$\hat{G}_j^*(r, L) = \left(\frac{1}{n_j}\right) \sum_j 1\{d_j \leq r\} \quad (4)$$

dengan d_j^* adalah jarak minimum dari titik uji ke titik latih yang tersedia dalam lipatan yang bersangkutan. Konfigurasi partisi dipilih berdasarkan nilai W terkecil dianggap paling mendekati kondisi prediksi sesungguhnya dan paling representatif untuk estimasi akurasi peta. Kedua metode partisi dibandingkan berdasarkan nilai W masing-masing terhadap acuan yang sama yaitu $\hat{G}_j(r)$. Dengan demikian nilai W dari kedua skema partisi tetap dapat diperbandingkan secara valid meskipun KFOLD menghasilkan lima subset melalui pembagian acak sedangkan BANDS menghasilkan dua subset melalui pembagian geografis, karena kedua nilai W tersebut mengukur hal yang sama yaitu seberapa jauh kondisi validasi masing-masing skema menyimpang dari kondisi prediksi, bukan seberapa mirip struktur kedua skema partisi itu satu sama lain.

Implementasi kNNDM dilakukan menggunakan fungsi `knndm` dari paket `CAST` versi 1.1.0 dalam R [45]. Jarak Euclidean dihitung pada sistem koordinat terproyeksi EPSG:32650, sehingga nilai statistik Wasserstein (W) memiliki satuan meter. Nilai W dikonversi dan disajikan dalam kilometer pada visualisasi ECDF. Titik prediksi diambil secara otomatis melalui pengambilan sampel beraturan sebanyak 1.000 titik dari domain model untuk tiap fase ENSO. Proses pembentukan k lipatan dilakukan melalui algoritma pengelompokan hierarkis dengan fungsi `penaut Ward` pada nilai default, dengan jumlah lipatan $k = 5$ yang disamakan dengan jumlah lipatan KFOLD manual agar perbandingan setara serta batas maksimum ukuran lipatan sebesar 0,5 yang berarti satu lipatan tidak diperkenankan memuat lebih dari 50% total titik latih.

2.7 Pemetaan Kerentanan dan Pemetaan Ekstrapolasi

Algoritma dengan nilai W terkecil diimplementasikan untuk membuat peta kerentanan kebakaran lahan gambut. Pembagian kategori (*breakpoint*) kerentanan menggunakan algoritma *Jenks natural breaks*, yaitu metode pengelompokan data yang meminimalkan variansi di dalam kelas sekaligus memaksimalkan variansi antarkelas [46]. *Breakpoint* dihitung secara terpisah untuk setiap fase ENSO dengan pembagian kelas sangat rendah, rendah, sedang, tinggi, dan sangat tinggi dengan persentase luasan provinsi. Pendekatan ini dipilih karena tidak ada satu algoritma unggul di semua kondisi [47]. Selain itu, skema pembelajaran setiap algoritma berbeda sehingga akan berpengaruh terhadap distribusi nilai kontinu kerentanan kebakaran lahan gambut di Kalimantan. Konsekuensi dari pendekatan ini yaitu tidak dapat dibandingkan secara seimbang untuk lintas fase, sehingga analisis lebih mengutamakan kerentanan untuk masing-masing fase dengan algoritma unggul yang berbeda. Selain itu disertakan kontribusi variabel prediktor untuk analisis lebih lanjut.

Ketika model tersebut diterapkan untuk seluruh kawasan gambut Kalimantan, terdapat kemungkinan model diminta mengekstrapolasi ke kombinasi nilai variabel yang tidak pernah ditemui selama proses kalibrasi. Analisis *Mobility-Oriented Parity* (MOP) berfungsi menilai keandalan ekstrapolasi tersebut dengan mengukur kemiripan ruang lingkungan antara area kalibrasi dan seluruh area proyeksi, terlepas dari algoritma pemodelan yang digunakan. Dengan kata lain, MOP tidak menilai performa model melainkan menilai sejauh mana wilayah yang diprediksi memiliki kombinasi kondisi lingkungan yang sebanding dengan wilayah tempat model belajar [48].

3. HASIL DAN PEMBAHASAN

3.1 Perbandingan Evaluasi Performa Algoritma Metode Partisi

Evaluasi algoritma dilakukan secara terpisah pada dua metode partisi untuk melihat sejauh mana pemilihan skema validasi memengaruhi estimasi akurasi model. Hasil evaluasi pada metode KFOLD disajikan pada Tabel 5, sedangkan hasil pada metode BANDS disajikan pada Tabel 6.

Tabel 5. Metrik Evaluasi Algoritma dengan Metode KFOLD

Fase ENSO	Metrik Evaluasi	SVM	BRT	RDF	GAM	GLM	MXD	MXS
El Niño	AUC	0,993	0,998	0,999	0,973	0,974	0,972	0,962
		± 0,001	± 0,001	± 0,000	± 0,001	± 0,001	± 0,001	± 0,001
El Niño	TSS	0,937	0,981	0,983	0,849	0,853	0,830	0,815
		± 0,004	± 0,002	± 0,003	± 0,003	± 0,003	± 0,003	± 0,006
Netral	AUC	0,991	0,996	0,998	0,976	0,979	0,966	0,958
		± 0,001	± 0,001	± 0,001	± 0,002	± 0,002	± 0,002	± 0,002
Netral	TSS	0,923	0,972	0,972	0,853	0,865	0,813	0,801
		± 0,005	± 0,004	± 0,003	± 0,002	± 0,005	± 0,010	± 0,006
La Niña	AUC	0,979	0,994	0,994	0,956	0,966	0,940	0,929
		± 0,007	± 0,005	± 0,003	± 0,012	± 0,010	± 0,018	± 0,012
La Niña	TSS	0,869	0,952	0,940 ±	0,838	0,873	0,748	0,752
		± 0,023	± 0,016	0,032	± 0,050	± 0,030	± 0,052	± 0,032

Pada metode KFOLD yang ditunjukkan Tabel 5, algoritma *Random Forest* (RDF) memperoleh nilai AUC tertinggi pada ketiga fase, yaitu 0,999 pada El Niño, 0,998 pada Netral, dan 0,994 pada La Niña. Untuk metrik TSS, RDF unggul pada El Niño dengan nilai 0,983, sejajar dengan *Boosted Regression Trees* (BRT) pada Netral dengan nilai 0,972, dan sedikit di bawah BRT pada La Niña 0,952. RDF dan BRT secara konsisten terpisah dari lima algoritma lain. GAM, GLM, MXD, dan MXS berada pada kisaran AUC 0,92 hingga 0,98 dengan TSS jauh lebih rendah, sekitar 0,75 hingga 0,87. Deviasi standar pada fase El Niño dan Netral untuk seluruh algoritma sangat kecil, tidak lebih dari 0,002 untuk AUC dan 0,010 untuk TSS. Pada fase La Niña, deviasi standar meningkat, mencapai 0,018 untuk AUC dan 0,052 untuk TSS pada MXD. Peningkatan ini sejalan dengan berkurangnya jumlah titik *hotspot* dari El Niño ke La Niña sebagaimana disajikan pada Tabel 3. Tingkat kestabilan yang tinggi pada partisi acak semacam ini umum terjadi ketika data latih dan data uji memiliki lokasi yang berdekatan secara spasial, sehingga performa model cenderung diperkirakan lebih tinggi dari kondisi generalisasi sesungguhnya [21], [49].

Tabel 6. Metrik Evaluasi Algoritma dengan Metode BANDS

Fase ENSO	Metrik Evaluasi	SVM	BRT	RDF	GAM	GLM	MXD	MXS
El Niño	AUC	0,877	0,885	0,875	0,715	0,725	0,929	0,795
		$\pm 0,024$	$\pm 0,108$	$\pm 0,106$	$\pm 0,288$	$\pm 0,273$	$\pm 0,021$	$\pm 0,173$
El Niño	TSS	0,654	0,687	0,631	0,457	0,475	0,703	0,511
		$\pm 0,025$	$\pm 0,172$	$\pm 0,214$	$\pm 0,364$	$\pm 0,339$	$\pm 0,029$	$\pm 0,272$
Netral	AUC	0,962	0,983	0,967	0,970	0,964	0,900	0,944
		$\pm 0,003$	$\pm 0,003$	$\pm 0,026$	$\pm 0,012$	$\pm 0,013$	$\pm 0,004$	$\pm 0,025$
Netral	TSS	0,801	0,888	0,815	0,842	0,839	0,701	0,769
		$\pm 0,007$	$\pm 0,007$	$\pm 0,097$	$\pm 0,046$	$\pm 0,032$	$\pm 0,025$	$\pm 0,068$
La Niña	AUC	0,783	0,840	0,870	0,736	0,698	0,694	0,776
		$\pm 0,079$	$\pm 0,080$	$\pm 0,021$	$\pm 0,019$	$\pm 0,099$	$\pm 0,124$	$\pm 0,001$
La Niña	TSS	0,514	0,595	0,662	0,410	0,370	0,419	0,500
		$\pm 0,102$	$\pm 0,150$	$\pm 0,081$	$\pm 0,021$	$\pm 0,161$	$\pm 0,165$	$\pm 0,049$

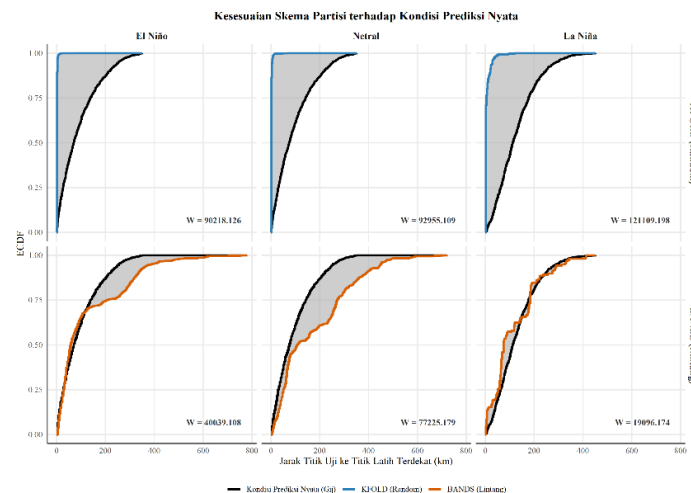
Performa keseluruhan pada metode BANDS turun signifikan dan deviasi standarnya melebar dibanding KFOLD, terutama pada fase El Niño dan La Niña pada Tabel 6. Algoritma dengan nilai tertinggi pada El Niño berpindah dari RDF ke *Maximum Entropy default* (MXD) dengan AUC 0,929 dan TSS 0,703, sementara RDF turun ke AUC 0,875 dengan deviasi standar melebar hingga 0,106 untuk AUC dan 0,214 untuk TSS. Penurunan performa RDF yang besar antara KFOLD dan BANDS ini menunjukkan bahwa nilai tinggi yang diperoleh RDF pada partisi acak sebagian besar disebabkan oleh kedekatan lokasi data latih dan data uji, bukan kemampuan generalisasi model yang sesungguhnya [19], [21]. Pada fase yang sama, GAM dan GLM memiliki deviasi standar yang lebih besar mencapai 0,288 dan 0,273 untuk AUC. Hal ini menunjukkan bahwa perbedaan komposisi data antarblok geografis lebih berperan dalam menentukan tingkat ketidakstabilan dibanding jenis algoritma yang digunakan [23]. Sebaliknya, MXD sebagai algoritma terpilih pada fase ini memiliki deviasi standar yaitu 0,021 untuk AUC dan 0,029 untuk TSS, sehingga MXD unggul pada nilai rata-rata dan menunjukkan hasil konsisten pada fase El Niño.

Pola berbeda terlihat pada fase Netral. Penurunan performa dari KFOLD ke BANDS lebih kecil, BRT hanya turun dari AUC 0,996 menjadi 0,983 dan TSS dari 0,972 menjadi 0,888. Deviasi standar BRT pada fase ini juga tetap kecil, yaitu 0,003 untuk AUC dan 0,007 untuk TSS, bahkan menjadi yang terkecil di antara seluruh algoritma pada fase Netral. GAM dan GLM pada fase ini turut menunjukkan deviasi standar yang relatif kecil, tidak lebih dari 0,046 untuk TSS, sehingga keduanya juga tergolong stabil. Meski demikian, nilai rata-rata BRT tetap lebih tinggi dibanding GAM dan GLM pada kedua metrik. Sehingga, pemilihan BRT sebagai algoritma terbaik pada fase Netral didasarkan pada keunggulan nilai rata-rata, bukan karena GAM dan GLM tidak stabil.

Pada fase La Niña, RDF menjadi algoritma dengan nilai tertinggi untuk kedua metrik di bawah BANDS, dengan AUC 0,870 dan TSS 0,662. Deviasi standar RDF pada fase ini, yaitu 0,021 untuk AUC dan 0,081 untuk TSS juga tergolong kecil dibanding algoritma lain pada fase yang sama. Nilai AUC dan TSS pada fase La Niña secara umum tetap lebih rendah dibanding dua fase lain, sejalan dengan jumlah titik *hotspot* yang hanya 314 setelah penjarangan spasial yang merupakan jumlah paling sedikit di antara ketiga fase. Keterbatasan jumlah data ini menjadi salah satu sumber ketidakpastian yang perlu dipertimbangkan dalam interpretasi hasil model La Niña [41]. Metode partisi acak dan spasial dapat menghasilkan parameter model optimal yang berbeda pada sebagian besar kasus yang diuji [50].

3.2 Perbandingan Metode Partisi Acak dengan Partisi Spasial Menggunakan kNNDM

Untuk menilai skema partisi yang lebih merepresentasikan kondisi prediksi sesungguhnya, dilakukan diagnostik kNNDM yang membandingkan ECDF jarak titik uji ke titik latih terdekat pada tiap metode partisi dalam fase ENSO yang berbeda terhadap ECDF kondisi prediksi nyata di seluruh domain gambut Kalimantan. Hasilnya disajikan pada Gambar 2.



Gambar 2. Perbandingan Kesesuaian Metode Partisi KFOLD dengan BANDS Berdasarkan Fungsi Distribusi Kumulatif Empiris Jarak Titik Uji ke Titik Latih Terdekat (kNNDM)

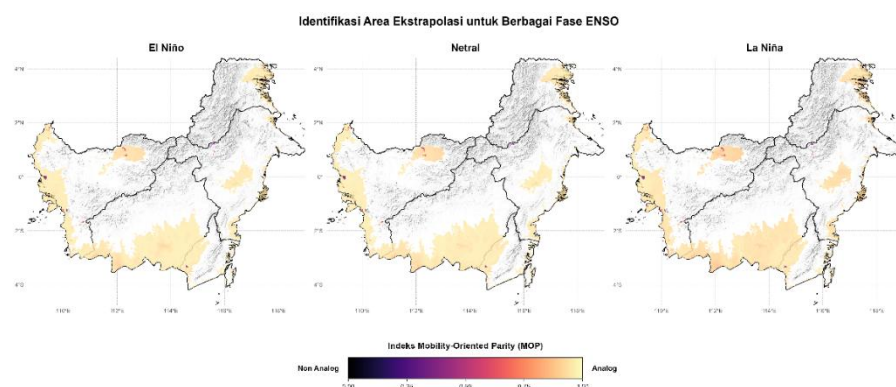
Baris atas pada Gambar 2 menunjukkan pemisahan yang jelas antara garis KFOLD dan garis kondisi prediksi nyata pada ketiga fase. Nilai statistik Wasserstein (W) mencetak nilai tertinggi di antara seluruh skema yang diuji, yaitu 90.218 meter pada El Niño, 92.955 meter pada Netral, dan 121.109 meter pada La Niña. Garis KFOLD yang menumpuk mendekati nol pada rentang jarak yang pendek menunjukkan bahwa sebagian besar titik uji KFOLD memiliki titik latih terdekat pada jarak yang jauh lebih dekat dibanding jarak yang sesungguhnya dihadapi model saat memprediksi ke seluruh domain. Performa tinggi KFOLD tidak merepresentasikan kondisi prediksi sesungguhnya [20].

Sementara itu, baris bawah menunjukkan penurunan nilai W yang cukup besar pada partisi BANDS di seluruh fase, yaitu 40.039 meter pada El Niño, 77.225 meter pada Netral, dan 19.096 meter pada La Niña. Penurunan pada fase La Niña sekitar seperenam nilai KFOLD pada fase yang sama, menunjukkan bahwa ECDF BANDS pada fase ini paling mendekati ECDF kondisi prediksi nyata. Fase Netral menunjukkan pola berbeda. Nilai W BANDS memang tetap lebih rendah dibanding KFOLD, namun penurunannya hanya sekitar 17%, yang lebih kecil dibanding penurunan sekitar 55% pada El Niño dan 84% pada La Niña.

Perbedaan besaran penurunan W antar fase ini sejalan dengan temuan bahwa keunggulan partisi spasial terhadap partisi acak tidak berlaku seragam, melainkan bergantung pada struktur data serta ukuran dan jumlah blok yang terbentuk [22], [23]. Fase La Niña, yang memiliki jumlah titik *hotspot* paling sedikit, justru menunjukkan perbaikan terbesar, sehingga blok geografis pada BANDS dapat menangkap struktur spasial data yang jarang dengan lebih baik dibanding partisi acak. Berdasarkan hasil kNNDM ini, algoritma unggul tiap fase pada Tabel 6 digunakan sebagai dasar pemetaan kerentanan kebakaran lahan gambut di Kalimantan, karena metode BANDS secara konsisten menghasilkan estimasi yang lebih mendekati kondisi prediksi sesungguhnya dibanding KFOLD meskipun nilai AUC dan TSS secara nominal lebih rendah.

3.3 Identifikasi Area Ekstrapolasi

Sebelum algoritma unggul tiap fase diterapkan pada seluruh domain gambut Kalimantan, keandalan ekstrapolasi model dinilai melalui analisis *Mobility-Oriented Parity* (MOP) yang ditunjukkan pada Gambar 3.



Gambar 3. Identifikasi Area Ekstrapolasi Pemodelan Berdasarkan Indeks *Mobility Oriented Parity* (MOP)

Sebagian besar wilayah gambut Kalimantan pada ketiga fase memiliki nilai MOP tinggi mendekati satu, terutama di gambut pesisir selatan Kalimantan Tengah dan Kalimantan Selatan serta pesisir timur Kalimantan

Timur ditampilkan pada Gambar 3. Kondisi ini menandakan bahwa kombinasi variabel lingkungan pada area tersebut umumnya berada dalam rentang yang telah dipelajari model selama kalibrasi. Area dengan nilai MOP rendah bersifat terbatas dan terlokalisasi, terkonsentrasi di perbatasan Kalimantan Barat dan Kalimantan Tengah di sekitar 112°BT, 0,5°LU, disertai beberapa titik kecil di sepanjang pesisir barat dan sebagian pesisir timur. Klaster tersebut tampak sedikit lebih luas pada fase La Niña, sebanding dengan jumlah titik *hotspot* La Niña yang jauh lebih sedikit, sehingga ruang lingkungan yang tercakup selama kalibrasi model relatif lebih sempit dibanding dua fase lainnya.

Area dengan nilai MOP rendah perlu diinterpretasikan secara hati-hati karena prediksi pada kombinasi lingkungan yang tidak terwakili selama kalibrasi berisiko menghasilkan estimasi yang kurang dapat diandalkan [33]. Meski demikian, karena proporsi area *non analog* pada penelitian ini relatif kecil dibanding total luas domain, prediksi kerentanan pada sebagian besar wilayah gambut Kalimantan dapat dianggap cukup andal. Analisis MOP pada bagian ini berfungsi sebagai identifikasi area yang memerlukan kehati-hatian interpretasi, bukan sebagai indikator akurasi model [39], [48].

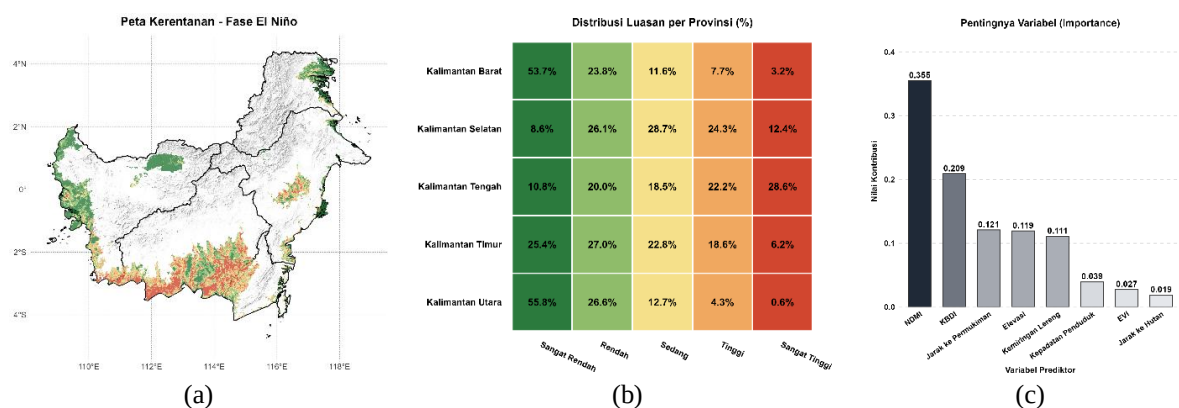
3.4 Pemetaan Kerentanan Kebakaran Lahan Gambut

Algoritma dengan performa terbaik pada tiap fase berdasarkan metode BANDS diterapkan untuk menghasilkan peta kerentanan kebakaran lahan gambut Kalimantan. Klasifikasi kelas kerentanan menggunakan *Jenks natural breaks* yang dihitung terpisah untuk tiap fase disajikan pada Tabel 7, sedangkan sebaran spasial, persentase luasan per provinsi, dan kontribusi variabel prediktor disajikan pada Gambar 4, Gambar 5, dan Gambar 6.

Tabel 7. Pembagian Klasifikasi Kelas Kerentanan Kebakaran Lahan Gambut

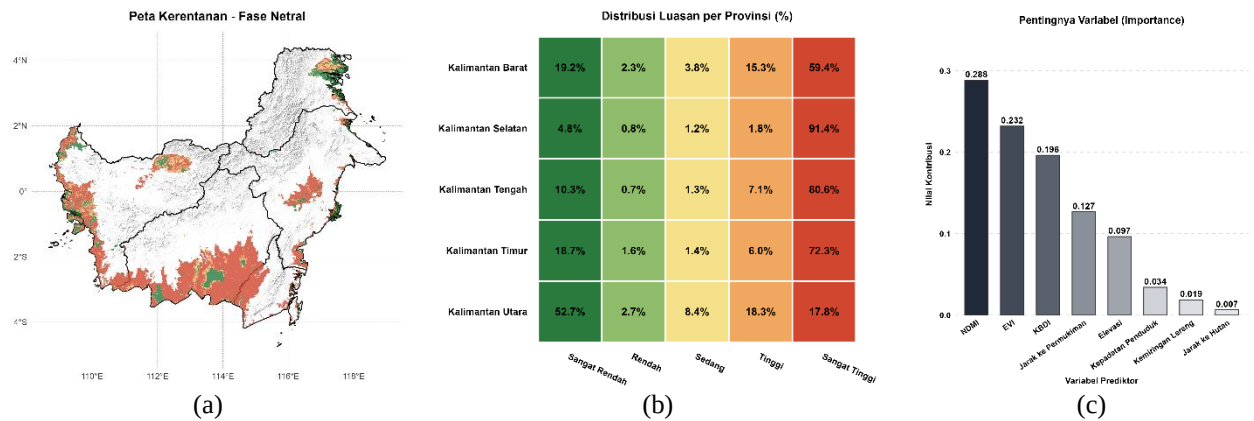
Fase ENSO	Algoritma	Nilai Minimum	Break 1	Break 2	Break 3	Break 4	Nilai Maksimum
El Niño	<i>Maximum Entropy default</i> (MXD)	0	0,1447	0,3061	0,4861	0,7043	1
Netral	<i>Boosted Regression Trees</i> (BRT)	0	0,1968	0,5444	0,8037	0,9482	1
La Niña	<i>Random Forest</i> (RDF)	0	0,1338	0,3656	0,6221	0,8423	1

Tabel 7 menunjukkan bahwa ketiga fase menggunakan algoritma dan titik potong kelas yang berbeda, yaitu MXD untuk El Niño, BRT untuk Netral, dan RDF untuk La Niña. Karena *breakpoint* dihitung terpisah untuk tiap fase, perbandingan proporsi luasan kelas kerentanan hanya berlaku dalam konteks masing-masing fase dan tidak dapat dibandingkan secara langsung untuk lintas fase. Ini berlaku karena tidak ada satu algoritma unggul di semua kondisi [47].



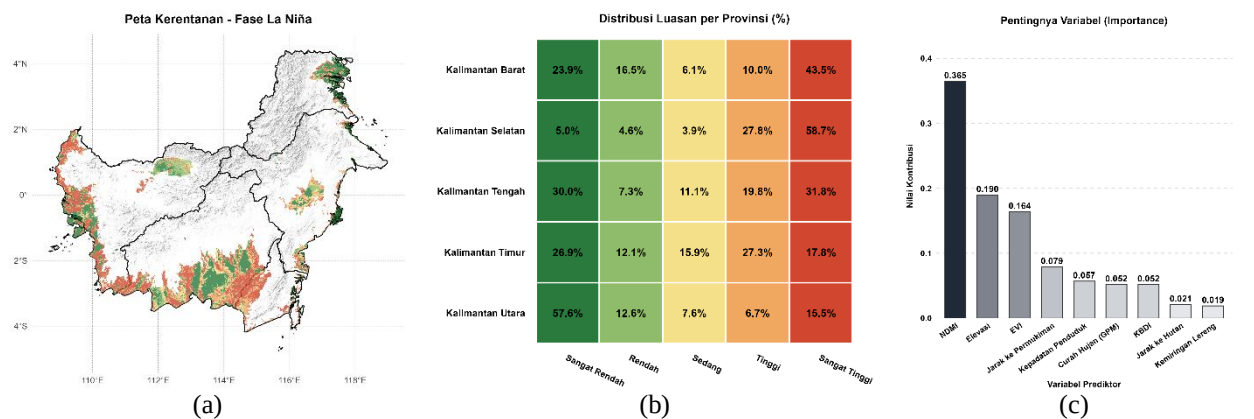
Gambar 4. Peta Distribusi Kelas Kerentanan Kebakaran Lahan Gambut Kalimantan pada Fase El Niño (a), Persentase Luasan Kelas Kerentanan pada Provinsi (b), dan Kontribusi Variabel Prediktor (c).

Pada Fase El Niño Kalimantan Tengah dan Kalimantan Selatan memiliki proporsi kelas tinggi dan sangat tinggi gabungan terbesar, masing-masing 50,8% dan 36,7% dari luas gambut provinsi yang ditunjukkan Gambar 4. Kalimantan Barat dan Kalimantan Utara didominasi kelas sangat rendah dan rendah, mencapai 77,5% dan 82,4%. Pola ini sesuai dengan catatan bahwa Kalimantan Tengah secara historis menjadi salah satu wilayah gambut paling rawan kebakaran di Indonesia [5], [27]. Pada bagian kontribusi variabel prediktor, NDMI menjadi yang paling dominan dengan nilai 0,355, diikuti KBDI sebesar 0,209, yang melampaui jarak ke permukiman, elevasi, dan kemiringan lereng. Dominasi NDMI dan KBDI menunjukkan bahwa kelembapan gambut dan tingkat kekeringan meteorologis menjadi pengendali utama kerentanan pada fase ini, sejalan dengan mekanisme penurunan muka air tanah yang meningkatkan paparan gambut terhadap oksigen [2], [3]. Hasil pemodelan kebakaran lahan gambut yang telah dilakukan di Sumatera, Kalimantan, dan Papua yang juga menempatkan indeks kelembapan dan jarak ke infrastruktur manusia sebagai prediktor utama [51].



Gambar 5. Peta Distribusi Kelas Kerentanan Kebakaran Lahan Gambut Kalimantan pada Fase Netral (a), Persentase Luasan Kelas Kerentanan pada Provinsi (b), dan Kontribusi Variabel Prediktor (c).

Pada Fase Netral yang ditunjukkan Gambar 5 kelas sangat tinggi mendominasi hingga 91,4% di Kalimantan Selatan, 80,6% di Kalimantan Tengah, dan 72,3% di Kalimantan Timur. Kalimantan Utara tetap didominasi kelas sangat rendah sebesar 52,7%. Kalimantan Barat justru menunjukkan pola berlawanan yaitu didominasi kelas sangat tinggi sebesar 59,4%. Urutan variabel prediktor pada fase ini bergeser, EVI naik ke peringkat kedua dengan nilai 0,232 menggantikan posisi KBDI yang turun ke peringkat ketiga dengan nilai 0,196, berbeda dari urutan pada El Niño. Pergeseran ini menunjukkan bahwa kondisi vegetasi turut berperan penting dalam menentukan kerentanan pada fase Netral [30].



Gambar 6. Peta Distribusi Kelas Kerentanan Kebakaran Lahan Gambut Kalimantan pada Fase La Niña (a), Persentase Luasan Kelas Kerentanan pada Provinsi (b), dan Kontribusi Variabel Prediktor (c).

Pada Fase La Niña yang ditampilkan Gambar 6, Kalimantan Selatan tetap memiliki proporsi kelas sangat tinggi terbesar yaitu 58,7%, namun Kalimantan Tengah menunjukkan distribusi yang lebih merata dengan kelas sangat rendah sebesar 30% dan sangat tinggi sebesar 31,8% yang hampir seimbang. Kalimantan Timur juga menunjukkan pola serupa, dengan kelas sangat rendah sebesar 26,9% hanya sedikit melampaui kelas tinggi dengan nilai 27,3%, berbeda dari dua fase lain yang cenderung terkonsentrasi pada kelas ekstrem. Kalimantan Barat menunjukkan proporsi kelas sangat tinggi sebesar 43,5%. Variabel NDMI tetap menjadi yang paling dominan dengan nilai 0,365, namun elevasi naik ke peringkat kedua dengan nilai 0,190, sementara KBDI turun ke peringkat keenam dengan nilai 0,052 yang sejajar dengan curah hujan. Penurunan peran KBDI pada fase La Niña sejalan dengan kondisi iklim yang lebih basah dan menghasilkan nilai indeks kekeringan lebih rendah. Nilai indeks yang seragam ini menurunkan nilai KBDI sebagai prediktor pada fase La Niña dibanding faktor topografi dan kelembapan vegetasi. Pola distribusi kelas yang lebih tinggi dari perkiraan pada beberapa provinsi ini perlu dibaca dengan mempertimbangkan keterbatasan jumlah data kalibrasi fase La Niña.

Secara keseluruhan, terdapat inkonsistensi hasil pemodelan yang tidak mencerminkan korelasi kebakaran dengan variasi ENSO. Beberapa kajian yang telah dilakukan menemukan bahwa pengaruh ENSO/IOD terhadap curah hujan relatif lemah di wilayah ekuatorial seperti Kalimantan Barat dan Sumatera Barat karena pola curah hujan lebih dikendalikan oleh konveksi lokal [52], [53]. Selain itu didapati anomali, yaitu peningkatan curah hujan dan banjir di Kalimantan Barat, Utara, dan Tengah saat El Niño kuat 2023, yang dikaitkan dengan pemanasan Laut China Selatan dan aktivitas awan lokal [54]. NDMI menjadi prediktor paling berpengaruh pada ketiga fase dengan nilai 0,355 pada El Niño, 0,288 pada Netral, dan 0,365 pada La Niña. Hasil ini menegaskan bahwa kondisi kelembapan gambut menjadi pengendali utama kerentanan kebakaran yang tidak bergantung pada fase ENSO, sejalan dengan peran drainase dan tutupan lahan sebagai determinan kejadian kebakaran gambut di Indonesia [8].

Prediktor sekunder yang mengikutinya berbeda antarfase, dari KBDI pada El Niño, EVI pada Netral, hingga elevasi pada La Niña, menunjukkan bahwa faktor pendukung kerentanan bersifat berbeda untuk tiap kondisi fase yang berbeda. Pola ini memperjelas alasan penggunaan algoritma dan klasifikasi yang terpisah untuk tiap fase, karena tidak ada satu algoritma yang unggul pada seluruh kondisi pemodelan distribusi [47].

4. KESIMPULAN

Penelitian ini membandingkan performa algoritma *machine learning* dengan dua metode partisi data dalam kerangka ENMTML untuk memetakan kerentanan kebakaran lahan gambut Kalimantan pada tiga fase ENSO periode 2015 – 2024. Metode partisi acak KFOLD secara konsisten menghasilkan nilai AUC dan TSS yang sangat tinggi, terutama pada *Random Forest* yang mencapai AUC 0,999 pada El Niño. Namun diagnostik kNNDM menunjukkan performa tersebut tidak mencerminkan kemampuan generalisasi sesungguhnya. Nilai statistik Wasserstein KFOLD berkisar 90.218 hingga 121.109 meter, jauh lebih tinggi dibanding BANDS yang hanya berkisar 19.096 hingga 77.225 meter. Metode BANDS karena itu dinilai lebih representatif terhadap kondisi prediksi meski nilai AUC dan TSS-nya lebih rendah secara nominal. Berdasarkan hasil BANDS, algoritma terbaik berbeda pada tiap fase ENSO. *Maximum Entropy default* unggul pada El Niño dengan AUC 0,929. *Boosted Regression Trees* unggul pada Netral dengan AUC 0,983. *Random Forest* unggul pada La Niña dengan AUC 0,870, lebih rendah dibanding dua fase lain karena jumlah titik *hotspot* yang hanya 314 setelah penjarangan spasial. Peta kerentanan menunjukkan pola spasial yang berbeda antarprovinsi dan antarfase. Perbedaan ini terkait dengan cara belajar algoritma dalam mengolah korelasi *hotspot* dengan variabel prediktor lingkungan, sehingga tidak dapat dibandingkan hasilnya dengan setara untuk tiap fase ENSO. Kalimantan Tengah dan Kalimantan Selatan memiliki proporsi kelas tinggi dan sangat tinggi dengan cakupan distribusi merata dan nilai tinggi dibandingkan Kalimantan Utara, Kalimantan Barat, dan Kalimantan Timur pada fase El Niño dan pada fase Netral. NDMI konsisten menjadi prediktor paling berpengaruh pada ketiga fase dengan nilai berkisar 0,288 sampai 0,365, menegaskan peran kelembapan gambut sebagai pengendali utama kerentanan kebakaran terlepas dari fase ENSO. Analisis MOP menunjukkan sebagian besar wilayah gambut Kalimantan berada dalam ruang lingkungan yang telah dipelajari model, dengan area ekstrapolasi rendah yang terbatas. Pengaruh variasi ENSO tidak menunjukkan kesesuaian dengan hasil pembelajaran algoritma, hal ini karena wilayah Kalimantan lebih dipengaruhi oleh aktivitas konveksi lokal. Penelitian ini menegaskan pentingnya validasi metode partisi data melalui diagnostik kNNDM sebelum menentukan skema evaluasi model kerentanan kebakaran berbasis spasial, karena partisi acak konvensional berisiko memberikan estimasi akurasi yang menyesatkan. Peta kerentanan kebakaran lahan gambut yang dihasilkan pada tiap fase ENSO ini diharapkan dapat menjadi dasar bagi perencanaan mitigasi dan konservasi lahan gambut yang lebih terarah dan sesuai dengan karakteristik lingkungan serta kondisi iklim di masing-masing provinsi di Kalimantan.

REFERENCES

- [1] S. L. Wilkinson, R. Andersen, P. A. Moore, S. J. Davidson, G. Granath, and J. M. Waddington, “Wildfire and degradation accelerate northern peatland carbon release,” *Nat. Clim. Chang.*, vol. 13, no. 5, pp. 456–461, May 2023, doi: 10.1038/s41558-023-01657-w.
- [2] L. Ma et al., “A globally robust relationship between water table decline, subsidence rate, and carbon release from peatlands,” *Commun Earth Environ*, vol. 3, no. 1, p. 254, Oct. 2022, doi: 10.1038/s43247-022-00590-8.
- [3] J. Koarashi et al., “Progressive release of long-stored carbon from tropical peatland disturbances,” *Nat Commun*, vol. 17, no. 1, p. 4369, May 2026, doi: 10.1038/s41467-026-72890-y.
- [4] L. Tacconi, “Preventing fires and haze in Southeast Asia,” *Nature Clim Change*, vol. 6, no. 7, pp. 640–643, Jul. 2016, doi: 10.1038/nclimate3008.
- [5] A. Schmidt et al., “Fire frequency, intensity, and burn severity in Kalimantan’s threatened Peatland areas over two Decades,” *Front. For. Glob. Change*, vol. 7, p. 1221797, Feb. 2024, doi: 10.3389/ffgc.2024.1221797.
- [6] W. Cai et al., “Changing El Niño–Southern Oscillation in a warming climate,” *Nat Rev Earth Environ*, vol. 2, no. 9, pp. 628–644, Aug. 2021, doi: 10.1038/s43017-021-00199-z.
- [7] S. Nurdianti, A. Sopaheluwakan, and P. Septiawan, “Joint Pattern Analysis of Forest Fire and Drought Indicators in Southeast Asia Associated with ENSO and IOD,” *Atmosphere*, vol. 13, no. 8, p. 1198, Jul. 2022, doi: 10.3390/atmos13081198.
- [8] R. Salmayenti, A. J. Baird, J. Holden, and D. V. Spracklen, “Drainage density and land cover interact to affect fire occurrence in Indonesian peatlands,” *Environ. Res. Lett.*, vol. 20, no. 5, p. 054036, May 2025, doi: 10.1088/1748-9326/adc755.
- [9] M. W. Jones et al., “Global and Regional Trends and Drivers of Fire Under Climate Change,” *Reviews of Geophysics*, vol. 60, no. 3, p. e2020RG000726, Sep. 2022, doi: 10.1029/2020RG000726.

- [10] W. Thuiller, “Ecological niche modelling,” *Current Biology*, vol. 34, no. 6, pp. R225–R229, Mar. 2024, doi: 10.1016/j.cub.2024.02.018.
- [11] N. Sillero et al., “Want to model a species niche? A step-by-step guideline on correlative ecological niche modelling,” *Ecological Modelling*, vol. 456, p. 109671, Sep. 2021, doi: 10.1016/j.ecolmodel.2021.109671.
- [12] S. Aldiansyah and F. Wardani, “Assessment of resampling methods on performance of landslide susceptibility predictions using machine learning in Kendari City, Indonesia,” *Water Practice & Technology*, vol. 19, no. 1, pp. 52–81, Jan. 2024, doi: 10.2166/wpt.2024.002.
- [13] D. Theng and K. K. Bhoyar, “Feature selection techniques for machine learning: a survey of more than two decades of research,” *Knowl Inf Syst*, vol. 66, no. 3, pp. 1575–1637, Mar. 2024, doi: 10.1007/s10115-023-02010-5.
- [14] L. Nguyen Van and G. Lee, “Optimizing Stacked Ensemble Machine Learning Models for Accurate Wildfire Severity Mapping,” *Remote Sensing*, vol. 17, no. 5, p. 854, Feb. 2025, doi: 10.3390/rs17050854.
- [15] C. Li et al., “Integrated spatial generalized additive modeling for forest fire prediction: a case study in Fujian Province, China,” *J. For. Res.*, vol. 36, no. 1, p. 30, Feb. 2025, doi: 10.1007/s11676-025-01822-1.
- [16] S. Janizadeh et al., “Combination four different ensemble algorithms with the generalized linear model (GLM) for predicting forest fire susceptibility,” *Geomatics, Natural Hazards and Risk*, vol. 14, no. 1, p. 2206512, Dec. 2023, doi: 10.1080/19475705.2023.2206512.
- [17] L. B. Prasetyo et al., “Assessing Sumatran Peat Vulnerability to Fire under Various Condition of ENSO Phases Using Machine Learning Approaches,” *Forests*, vol. 13, no. 6, p. 828, May 2022, doi: 10.3390/f13060828.
- [18] A. F. A. D. Andrade, S. J. E. Velazco, and P. De Marco Júnior, “ENMTML: An R package for a straightforward construction of complex ecological niche models,” *Environmental Modelling & Software*, vol. 125, p. 104615, Mar. 2020, doi: 10.1016/j.envsoft.2019.104615.
- [19] V. Bahn and B. J. McGill, “Testing the predictive performance of distribution models,” *Oikos*, vol. 122, no. 3, pp. 321–331, Mar. 2013, doi: 10.1111/j.1600-0706.2012.00299.x.
- [20] R. Valavi, J. Elith, J. J. Lahoz-Monfort, and G. Guillera-Aroita, “block CV : An r package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models,” *Methods Ecol Evol*, vol. 10, no. 2, pp. 225–232, Feb. 2019, doi: 10.1111/2041-210X.13107.
- [21] N. Karasiak, J.-F. Dejoux, C. Monteil, and D. Sheeren, “Spatial dependence between training and test sets: another pitfall of classification accuracy assessment in remote sensing,” *Mach Learn*, vol. 111, no. 7, pp. 2715–2740, Jul. 2022, doi: 10.1007/s10994-021-05972-1.
- [22] R. Valavi, J. Elith, J. J. Lahoz-Monfort, and G. Guillera-Aroita, “Flexible species distribution modelling methods perform well on spatially separated testing data,” *Global Ecol Biogeogr*, vol. 32, no. 3, pp. 369–383, Mar. 2023, doi: 10.1111/geb.13639.
- [23] A. Stock, “Choosing blocks for spatial cross-validation: lessons from a marine remote sensing case study,” *Front. Remote Sens.*, vol. 6, p. 1531097, Mar. 2025, doi: 10.3389/frsen.2025.1531097.
- [24] J. Linnenbrink, C. Milà, M. Ludwig, and H. Meyer, “kNNDM CV: k-fold nearest-neighbour distance matching cross-validation for map accuracy estimation,” *Geosci. Model Dev.*, vol. 17, no. 15, pp. 5897–5912, Aug. 2024, doi: 10.5194/gmd-17-5897-2024.
- [25] C. Milà, J. Mateu, E. Pebesma, and H. Meyer, “Nearest neighbour distance matching Leave-One-Out Cross-Validation for map validation,” *Methods Ecol Evol*, vol. 13, no. 6, pp. 1304–1316, Jun. 2022, doi: 10.1111/2041-210X.13851.
- [26] J. Linnenbrink, J. Nowosad, and H. Meyer, “Moving beyond spatial and random cross-validation in environmental modelling: a call for prediction-domain adaptive evaluation,” 2026, arXiv. doi: 10.48550/ARXIV.2605.13689.
- [27] A. J. Horton, J. Lehtinen, and M. Kumm, “Targeted land management strategies could halve peatland fire occurrences in Central Kalimantan, Indonesia,” *Commun Earth Environ*, vol. 3, no. 1, p. 204, Sep. 2022, doi: 10.1038/s43247-022-00534-2.
- [28] M. Unik, I. Sukaesih Sitanggang, L. Syaufina, and I. N. Surati Jaya, “Application of Random Forest Algorithm to Analyze the Confidence Level of Forest Fire Hotspots in Riau Peatland,” *JPSL*, vol. 15, no. 2, p. 255, Mar. 2025, doi: 10.29244/jpsl.15.2.255.
- [29] L. Del Rio Amador, M. Boudreault, and D. A. Carozza, “Global Asymmetries in the Influence of ENSO on Flood Risk Based on 1,600 Years of Hybrid Simulations,” *Geophysical Research Letters*, vol. 50, no. 5, p. e2022GL102027, Mar. 2023, doi: 10.1029/2022GL102027.
- [30] F. Ren, N. Tobinsky, T. Dorji, and A. Guibaud, “On the importance of both climate and vegetation evolution when predicting long-term wildfire susceptibility,” *International Journal of Wildland Fire*, vol. 34, no. 12, p. WF25092, Dec. 2025, doi: 10.1071/WF25092.
- [31] J. Schlör, F. Strnad, A. Capotondi, and B. Goswami, “Contribution of El Niño Southern Oscillation (ENSO) Diversity to Low-Frequency Changes in ENSO Variance,” *Geophysical Research Letters*, vol. 51, no. 14, p. e2024GL109179, Jul. 2024, doi: 10.1029/2024GL109179.
- [32] N. Phelps and D. G. Woolford, “Comparing calibrated statistical and machine learning methods for wildland fire occurrence prediction: a case study of human-caused fires in Lac La Biche, Alberta, Canada,” *International Journal of Wildland Fire*, vol. 30, no. 11, pp. 850–870, Nov. 2021, doi: 10.1071/WF20139.
- [33] S. J. E. Velazco, F. Villalobos, F. Galvão, and P. De Marco Júnior, “A dark scenario for Cerrado plant species: Effects of future climate, land use and protected areas ineffectiveness,” *Diversity and Distributions*, vol. 25, no. 4, pp. 660–673, Apr. 2019, doi: 10.1111/ddi.12886.
- [34] Q. Lamboley and Y. Fourcade, “No optimal spatial filtering distance for mitigating sampling bias in ecological niche models,” *Journal of Biogeography*, vol. 51, no. 9, pp. 1783–1794, Sep. 2024, doi: 10.1111/jbi.14854.

- [35] C. Ten Caten and T. Dallas, “Thinning occurrence points does not improve species distribution model performance,” *Ecosphere*, vol. 14, no. 12, p. e4703, Dec. 2023, doi: 10.1002/ecs2.4703.
- [36] M. H. Saputra et al., “Tracking Sumatran Tiger (*Panthera tigris sumatrae* Pocock, 1929) distribution in Gunung Leuser National Park: The influence of prey presence and environmental variables on habitat selection,” *NC*, vol. 59, pp. 43–64, Jul. 2025, doi: 10.3897/natureconservation.59.144262.
- [37] A. Rausell-Moreno, N. Galiana, B. Naimi, and M. B. Araújo, “Improving species distribution models by optimising background points: Impacts on current and future climate projections,” *Ecological Modelling*, vol. 507, p. 111177, Aug. 2025, doi: 10.1016/j.ecolmodel.2025.111177.
- [38] Fauziah, L. B. Prasetyo, N. Saribanon, and N. Hayati, “Vulnerability of peatland fires in bengkalis regency during the ENSO El nino phase using a machine learning approach,” *MethodsX*, vol. 14, p. 103128, Jun. 2025, doi: 10.1016/j.mex.2024.103128.
- [39] M. E. Cobos, H. L. Owens, J. Soberón, and A. T. Peterson, “Detailed multivariate comparisons of environments with mobility oriented parity,” *biogeography*, vol. 17, p. e132916, Sep. 2024, doi: 10.21425/fob.17.132916.
- [40] G. Vos, L. Van Eijk, Z. Sarnyai, and M. Rahimi Azghadi, “Stabilizing machine learning for reproducible and explainable results: A novel validation approach to subject-specific insights,” *Computer Methods and Programs in Biomedicine*, vol. 269, p. 108899, Sep. 2025, doi: 10.1016/j.cmpb.2025.108899.
- [41] T. Abedin, H. Xu, and S. Uddin, “The impact of K selection in K-fold cross-validation on bias and variance in supervised learning models,” *Sci Rep*, vol. 16, no. 1, p. 6084, Jan. 2026, doi: 10.1038/s41598-026-37247-x.
- [42] S. Lee et al., “Measures to improve the accuracy of species distribution models for effective conservation of endangered species,” *Global Ecology and Conservation*, vol. 66, p. e04061, Apr. 2026, doi: 10.1016/j.gecco.2026.e04061.
- [43] C. Shi and F. Zhang, “A Forest Fire Susceptibility Modeling Approach Based on Integration Machine Learning Algorithm,” *Forests*, vol. 14, no. 7, p. 1506, Jul. 2023, doi: 10.3390/f14071506.
- [44] O. Allouche, A. Tsoar, and R. Kadmon, “Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS),” *Journal of Applied Ecology*, vol. 43, no. 6, pp. 1223–1232, Dec. 2006, doi: 10.1111/j.1365-2664.2006.01214.x.
- [45] H. Meyer, M. Ludwig, C. Milà, J. Linnenbrink, and F. Schumacher, “The CAST Package for Training and Assessment of Spatial Prediction Models,” in *R Coding for Ecology*, D. Rocchini, Ed., in *Use R!*, Cham: Springer Nature Switzerland, 2026, pp. 247–266. doi: 10.1007/978-3-031-99665-8_11.
- [46] Z. Özcan, İ. Caglayan, Ö. Kabak, and F. Kılıç Gül, “Integrated risk mapping for forest fire management using the analytical hierarchy process and ordered weighted average: a case study in southern Turkey,” *Nat Hazards*, vol. 121, no. 1, pp. 959–1001, Jan. 2025, doi: 10.1007/s11069-024-06810-y.
- [47] J. Montoya-Jiménez, J. Valdez-Lazalde, G. Ángeles-Perez, H. De Los Santos-Posadas, and G. Cruz-Cárdenas, “Predictive capacity of nine algorithms and an ensemble model to determine the geographic distribution of tree species,” *iForest*, vol. 15, no. 5, pp. 363–371, Oct. 2022, doi: 10.3832/for4084-015.
- [48] Syartinilia, A. A. Condro, and S. Tsuyuki, “Projected impacts of climate change and anthropogenic effects on habitat distribution of endangered Javan Hawk-Eagle in Indonesia,” *Geography and Sustainability*, vol. 5, no. 2, pp. 241–250, Jun. 2024, doi: 10.1016/j.geosus.2024.01.009.
- [49] T. Kattenborn, F. Schiefer, J. Frey, H. Feilhauer, M. D. Mahecha, and C. F. Dormann, “Spatially autocorrelated training and validation samples inflate performance assessment of convolutional neural networks,” *ISPRS Open Journal of Photogrammetry and Remote Sensing*, vol. 5, p. 100018, Aug. 2022, doi: 10.1016/j.ojphoto.2022.100018.
- [50] H. Huang et al., “Cross-validation matters in species distribution models: a case study with goatfish species,” *Ecography*, vol. 2025, no. 1, p. e07354, Jan. 2025, doi: 10.1111/ecog.07354.
- [51] W. S. Karurung, K. Lee, and W. Lee, “Assessment of forest fire vulnerability prediction in Indonesia: Seasonal variability analysis using machine learning techniques,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 138, p. 104435, Apr. 2025, doi: 10.1016/j.jag.2025.104435.
- [52] Y. P. Dewanti, M. Muliadi, and R. Adriat, “Pengaruh El Niño Southern Oscillation (ENSO) Terhadap Curah Hujan di Kalimantan Barat,” *Prism. Fis.*, vol. 6, no. 3, Oct. 2018, doi: 10.26418/pf.v6i3.28699.
- [53] M. Ariska, A. Darmawan, S. Supari, M. Irfan, and I. Iskandar, “Analysis of the impact climate anomalies (ENSO and IOD) on environments based of computing in the Western Sumatra Region (Equatorial Region of Indonesia),” *J. Aceh Phys. Soc.*, vol. 12, no. 2, pp. 12–18, Apr. 2023, doi: 10.24815/jacps.v12i2.31167.
- [54] S. Zehri et al., “Diverse impact of 2023 El Niño on weather patterns over the Indonesian Maritime Continent,” *Journal of Southern Hemisphere Earth Systems Science*, vol. 75, no. 2, p. ES25005, Aug. 2025, doi: 10.1071/ES25005.