PEMETAAN KERAWANAN LONGSOR DAN KETIDAKPASTIAN PREDIKSI BERBASIS EXPLAINABLE DEEP LEARNING MENGGUNAKAN SENTINEL-1 DAN TURUNAN TOPOGRAFI DI SULAWESI TENGGARA

Cakra[*1], Andi Muhammad Islah[1], Baharuddin Rahman[1], Asniar[2]
[1]Program Studi Sistem Komputer, Program Studi Komputerisasi Akuntasi, STMIK Catur Sakti Kendari, Indonesia
[2]Program Studi Komputerisasi Akuntansi, Fakultas Ilmu Terapan, Universitas Telkom, Bandung, Indonesia
Penulis korespondesi: Cakra (ctjantong@gmail.com)

Tanah longsor membutuhkan pemetaan kerawanan yang mempertimbangkan validitas geografis, keterlacakan label, dan ketidakpastian prediksi. Penelitian ini mengembangkan kerangka Explainable Deep Learning berbasis citra Sentinel-1 GRD dan turunan topografi di Sulawesi Tenggara. Sebanyak 14 tile dimosaikkan menjadi raster 10 kanal untuk mengekstraksi 54 patch berukuran 128 × 128 piksel. Pembagian berbasis blok spasial menghasilkan 36 data latih, 9 validasi, dan 9 uji; normalisasi p01–p99 dihitung hanya dari data latih. Light CNN menghasilkan accuracy 0,1111 dan macro AUC-ROC 0,3148, menunjukkan kemampuan diskriminasi yang terbatas. Pemeriksaan geografis terhadap 500 pusat grid mengindikasikan 189 titik di daratan, 311 di perairan, dan 59 patch diperkirakan beririsan dengan pantai. Seluruh titik perairan diprediksi sebagai kelas Sedang, sehingga dominasi kelas tersebut belum merepresentasikan kerawanan daratan. Grad-CAM, SHAP, Monte Carlo Dropout, dan Deep Ensemble digunakan untuk menelaah perilaku serta ketidakpastian model. Kontribusi penelitian terletak pada kerangka evaluasi geospasial yang transparan dan dapat direproduksi; validasi sumber label dan inferensi berbatas daratan tetap diperlukan. Dengan demikian, hasil penelitian menjadi dasar pengembangan peta yang lebih andal setelah verifikasi spasial dan label.

Kata Kunci — Tanah Longsor, Explainable Deep Learning, Sentinel-1 GRD, Ketidakpastian Prediksi, Validitas Geospasial.

I. PENDAHULUAN

Tanah longsor merupakan proses geomorfologi yang dipengaruhi interaksi antara kemiringan lereng, kondisi material, air, tutupan lahan, dan aktivitas manusia. Kualitas inventaris kejadian atau kelas kerawanan menjadi fondasi penting karena kesalahan posisi, ketidakseimbangan sebaran, dan ketidakseragaman label dapat langsung memengaruhi kemampuan model dalam mengenali hubungan antara kondisi medan dan tingkat kerawanan [1].

Penelitian terdahulu oleh tim pengusul telah menunjukkan bahwa integrasi Sentinel-1 dan variabel topografi dapat digunakan untuk membangun model CNN tiga kelas pada Sulawesi Tenggara. Meskipun demikian, keluaran model sebelumnya masih berfokus pada label prediksi dan belum memberi informasi yang memadai tentang alasan spasial, kontribusi fitur, serta seberapa yakin model terhadap setiap keputusan [2].

Sentinel-1 memiliki keunggulan untuk wilayah tropis karena sistem radar dapat mengamati permukaan tanpa bergantung pada pencahayaan dan relatif tidak terhambat awan. Produk GRD dapat diproses melalui kalibrasi radiometrik, reduksi speckle, dan koreksi medan sebelum dikombinasikan dengan DEM. Pendekatan tersebut memungkinkan pembentukan fitur radar dan topografi yang konsisten untuk analisis citra multikanal [3], [4].

Jaringan saraf konvolusional mampu mempelajari struktur lokal, tekstur, dan pola spasial secara hierarkis. Akan tetapi, peningkatan kapasitas model tidak selalu diikuti peningkatan generalisasi, khususnya ketika ukuran dataset terbatas. Karena itu, performa klasifikasi perlu dianalisis bersama perilaku pelatihan, kesalahan per kelas, interpretabilitas, dan ketidakpastian probabilistik [5].

Kebaruan penelitian ini terletak pada penyatuan empat keluaran dalam satu alur: klasifikasi kerawanan, penjelasan spasial melalui Grad-CAM, pemeringkatan kontribusi fitur melalui SHAP, dan pengukuran ketidakpastian melalui Monte Carlo Dropout serta Deep

Ensemble. Pendekatan ini relevan untuk konteks mitigasi di Sulawesi Tenggara karena hasil prediksi dapat dibaca bersama indikator keyakinan, sehingga lokasi dengan keluaran tidak stabil dapat diprioritaskan untuk verifikasi lapangan [6].

Penelitian bertujuan:

1. Membangun dan mengevaluasi Light CNN berbasis patch multikanal untuk klasifikasi kerawanan rendah, sedang, dan tinggi setelah koreksi spatial leakage dan normalisasi training-only.
2. Menganalisis perilaku model melalui Grad-CAM dan SHAP secara diagnostik dengan mempertimbangkan kemampuan diskriminasi model pada data uji independen.
3. Mengestimasi ketidakpastian prediksi menggunakan Monte Carlo Dropout dan Deep Ensemble serta membandingkan confidence dan entropy pada data uji.
4. Menghasilkan keluaran spasial eksperimental berupa kelas prediksi, probabilitas kelas, confidence, dan predictive entropy pada pusat patch grid menggunakan statistik normalisasi dari data latih.

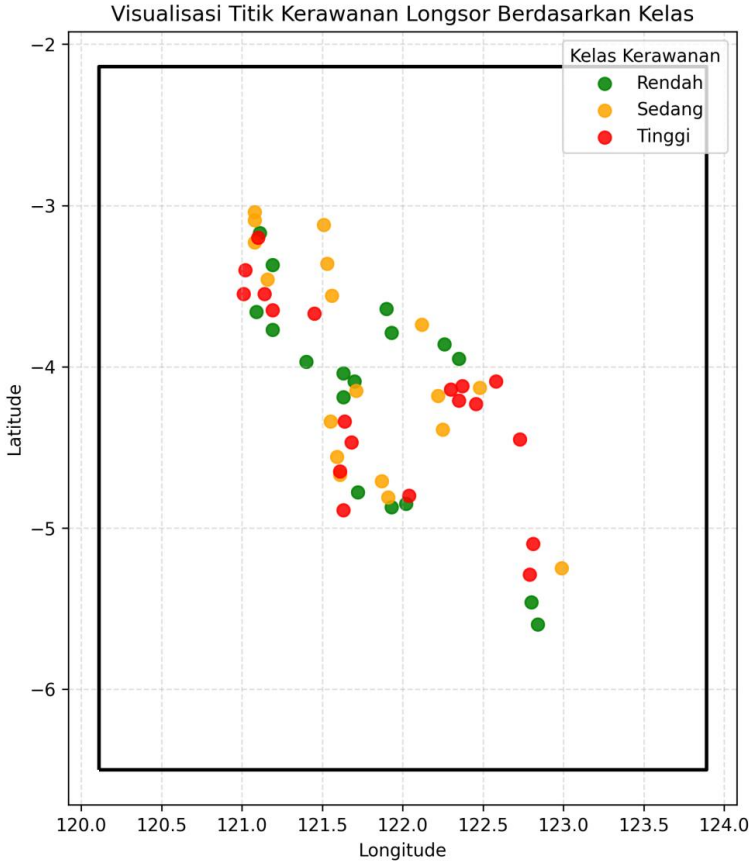
II. METODE PENELITIAN

A. Wilayah Kajian dan Sumber Data

Wilayah analisis berada di Sulawesi Tenggara. Dataset kerja memuat 54 catatan berkoordinat pada Kolaka, Kolaka Utara, Konawe, Konawe Selatan, Bombana, dan Buton, dengan label awal 17 Rendah, 18 Sedang, serta 19 Tinggi. Pemeriksaan notebook memperlihatkan bahwa koordinat, nama lokasi, dan label dimasukkan melalui daftar Python lalu ditulis menjadi CSV. Walaupun bagian kode menyebut “InaRisk BNPB”, tidak ditemukan rujukan file unduhan resmi, layer/versi, waktu pengambilan, pengenal fitur, maupun prosedur pengaitan label pada setiap titik. Dengan demikian, sumber primer dan validitas ground truth 54 catatan tersebut belum dapat dipastikan hanya dari naskah dan notebook yang tersedia.

InaRISK merupakan platform BNPB untuk informasi bahaya, kerentanan, kapasitas, dan risiko; ketiganya tidak dapat dipertukarkan. Ambang nilai rendah–sedang–tinggi harus sesuai dengan layer dan versi indikator yang benar-benar digunakan, bukan ditetapkan dari jumlah sampel atau diasumsikan dari kelas kabupaten. Jika label tidak memiliki bukti sumber, eksperimen supervised learning perlu dinyatakan bersifat eksploratif dan diulang setelah kurasi label independen [7].

Data acuan harus diaudit terhadap ekspor GIS, dokumen survei, atau dokumen resmi penyedia, dengan metadata minimum: instansi/URL dan nama layer, tanggal pengambilan, sistem koordinat, akurasi posisi, metode pengambilan titik, asal kelas, nilai indeks asal, ambang kelas, serta bukti validasi. Status titik juga harus dibedakan antara kejadian longsor aktual, titik survei bahaya/kerawanan, atau sampel yang diberi label melalui ekstraksi peta bahaya. Sebelum bukti tersedia, istilah yang digunakan adalah “titik referensi berlabel awal”, bukan “54 kejadian longsor terverifikasi”.



Gambar 1. Sebaran 54 titik inventaris kerawanan longsor berdasarkan koordinat geografis dan kelas kerawanan

Gambar 1 memperlihatkan posisi 54 titik referensi dan label awal Rendah, Sedang, serta Tinggi. Gambar ini menunjukkan sebaran koordinat pada dataset kerja, bukan lokasi kejadian longsor terverifikasi. Konsistensi koordinat dengan polygon daratan, kelayakan semantik label, dan dokumen sumber per titik harus diperiksa secara terpisah sebelum titik diperlakukan sebagai ground truth klasifikasi.

Tabel 1. Distribusi inventaris kerawanan longsor

Kelas	Jumlah	Persentase
Rendah	17	31,48%
Sedang	18	33,33%
Tinggi	19	35,19%

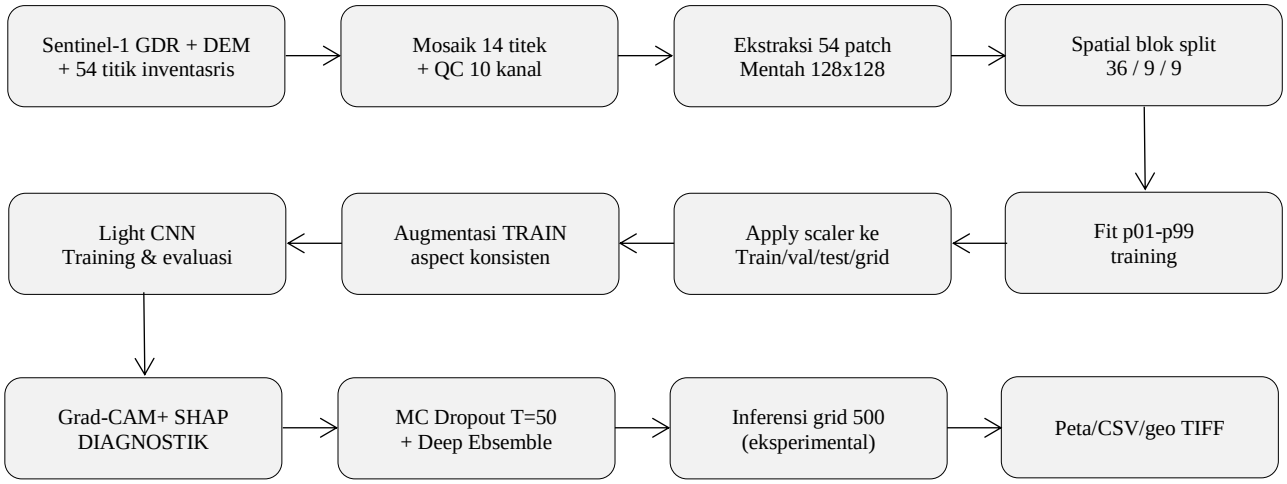
Tabel 1 merangkum distribusi label awal pada CSV kerja: 17 Rendah (31,48%), 18 Sedang (33,33%), dan 19 Tinggi (35,19%). Distribusi ini menggambarkan komposisi dataset yang telah digunakan dalam eksperimen, bukan estimasi proporsi kelas bahaya atau risiko di seluruh provinsi.

Keseimbangan jumlah label tidak membuktikan ketepatan kategorisasi. Validitas supervised learning bergantung pada kesesuaian definisi target (kejadian, bahaya, kerawanan, atau risiko), ketepatan koordinat, independensi anotasi, serta metode penentuan ambang setiap kelas. Karena dokumen sumber per titik belum tersedia, seluruh evaluasi klasifikasi harus dibaca sebagai hasil bersyarat pada label awal.

Titik referensi digunakan sebagai pusat ekstraksi patch radar dan topografi. Audit pendahuluan dengan garis pantai GSHHS resolusi menengah juga menandai satu koordinat pada daftar kerja, yaitu titik bernama “Latambaga” (121,63° BT; −4,89° LS), sebagai berpotensi berada di perairan. Temuan ini belum merupakan koreksi koordinat definitif dan harus diverifikasi terhadap layer daratan yang lebih rinci serta dokumen sumber. Pemeriksaan label dan lokasi merupakan prasyarat sebelum evaluasi ulang model.

B. Diagram Alur Penelitian

Penelitian mencakup pemeriksaan provenance dan label acuan, pengumpulan citra Sentinel-1/DEM, mosaik serta kontrol kualitas raster, ekstraksi patch mentah, pemisahan data berbasis blok spasial, fitting normalisasi hanya pada data latih, pelatihan dan evaluasi Light CNN, Grad-CAM/SHAP diagnostik, serta estimasi ketidakpastian. Tahap inferensi spasial yang layak untuk pelaporan ditambah validasi polygon daratan, klasifikasi titik darat–laut–pesisir, pemeriksaan mask NoData tiap kanal, dan rerun prediksi hanya pada patch daratan yang lolos seluruh pemeriksaan.



Gambar 2. Alur penelitian berbasis klasifikasi, interpretabilitas, ketidakpastian, dan inferensi spasial, dengan mosaik raster dan kontrol kualitas sebelum ekstraksi patch serta normalisasi p01–p99 berbasis training set

Gambar 2 menggambarkan urutan pemrosesan utama: mosaik dan validasi raster dilaksanakan sebelum ekstraksi patch; pemisahan spasial dilaksanakan sebelum perhitungan p01–p99, dengan statistik hanya dari data latih. Untuk revisi geospasial lanjutan, diperlukan land mask berbasis polygon ber-CRS konsisten, pemeriksaan footprint patch terhadap garis pantai, dan pengujian validitas piksel raster sebelum normalisasi maupun inferensi. Transformasi Aspect_sin/Aspect_cos harus tetap konsisten saat augmentasi. Data acuan dan cara penentuan kelas harus diverifikasi sebelum model dilatih ulang.

C. Pra-pemrosesan dan Raster Stack

Citra Sentinel-1 GRD periode 1 Januari–31 Desember 2025 difilter pada mode IW dan polarisasi VV/VH. Pada AOI final berbasis 54 titik inventaris dengan buffer 5 km diperoleh 266 scene, kemudian komposit median dibentuk. VV dan VH dikonversi ke skala linear untuk menghitung RVI, NDPI, dan CR. DEM SRTM digunakan untuk Elevation, Slope, Aspect, dan Curvature. Aspect tidak digunakan sebagai derajat kontinu, tetapi diubah menjadi Aspect_sin dan Aspect_cos agar sifat sirkular arah lereng terwakili. Stack 10 kanal diekspor menjadi 14 GeoTIFF, dimosaikkan dengan GDAL, dan menghasilkan raster final EPSG:4326 berukuran 10×9834×7681. Normalisasi robust mengikuti $x' = \text{clip}((x - P01_{train})/(P99_{train} - P01_{train}), 0, 1)$, dengan P01_train dan P99_train dihitung hanya dari data training.

$RVI = 4VH/(VV + VH)$; $NDPI = (VV - VH)/(VV + VH)$; $CR = VH/VV$; $Aspect_sin = \sin(Aspect \times \pi/180)$; $Aspect_cos = \cos(Aspect \times \pi/180)$.

Tabel 2. Sepuluh kanal masukan model setelah transformasi Aspect

Sumber: Hasil pembentukan raster stack penelitian, 2026.

Tabel 2. Sepuluh kanal masukan model setelah transformasi Aspect
Sumber: Hasil pembentukan raster stack penelitian, 2026.

Kanal	Sumber	Fungsi ringkas
VV	Sentinel-1	Hamburan permukaan
VH	Sentinel-1	Vegetasi/struktur volume
RVI	Turunan radar	Variasi vegetasi
NDPI	Turunan radar	Perbedaan polarisasi
CR	Turunan radar	Rasio VH terhadap VV
Elevation	DEM	Ketinggian medan
Slope	DEM	Kemiringan lereng
Aspect_sin	DEM	Komponen sinus arah lereng
Aspect_cos	DEM	Komponen cosinus arah lereng
Curvature	DEM	Bentuk/kelengkungan medan

D. Ekstraksi Patch dan Pembagian Data

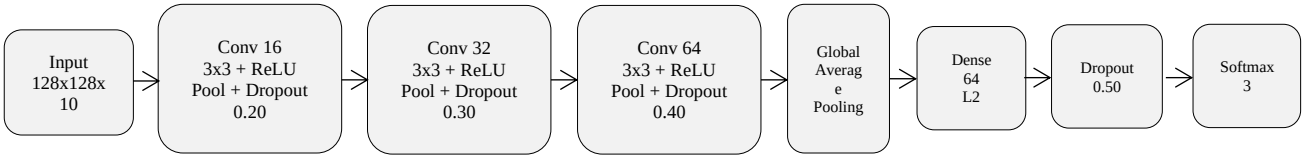
Patch awal berukuran 128×128×10 berhasil diekstraksi pada 54 koordinat dataset kerja. Pengujian sebelumnya memperoleh valid_fraction dan finite_fraction 1,00 serta spatial split 36/9/9 tanpa overlap footprint antarsubset. Akan tetapi, pemeriksaan numerik terhadap NaN, NoData, dan ketercakupan raster tidak secara otomatis membuktikan bahwa seluruh pixel pada patch berada di daratan ataupun bahwa labelnya sah. Oleh karena itu, kontrol kualitas akan diperluas dengan pemeriksaan polygon daratan, mask/NoData asli per kanal, footprint dalam area analisis, serta audit provenance 54 label sebelum semua metrik dilaporkan kembali sebagai evaluasi final.

Tabel 3. Pembagian dataset setelah mosaik penuh, kontrol kualitas, dan spatial split (hasil rerun)

Subset	Jumlah	Rendah	Sedang	Tinggi
Latih	36	10	13	13
Validasi	9	4	2	3
Uji	9	3	3	3

E. Arsitektur dan Pelatihan Model

Model final menggunakan Light CNN dengan input 128×128×10. Tiga blok konvolusi menggunakan 16, 32, dan 64 filter 3×3, masing-masing diikuti ReLU, max pooling, serta dropout 0,20; 0,30; dan 0,40. Representasi spasial dipadatkan dengan Global Average Pooling, Dense 64 berregularisasi L2, dropout 0,50, dan softmax tiga kelas. Total parameter model adalah 28.947. Arsitektur ini sengaja dipertahankan ringan mengingat ukuran dataset terbatas.



Gambar 3. Arsitektur Light CNN Pada Pipeline

Gambar 3. Arsitektur Light CNN pada pipeline dengan input 128×128×10 dan total 28.947 parameter. Pelatihan menggunakan Adam dengan learning rate awal 0,0003, categorical cross-entropy dengan label smoothing 0,05, batch size 8, dan maksimum 100 epoch. ModelCheckpoint serta EarlyStopping memantau validation AUC, sedangkan ReduceLROnPlateau memantau validation loss. Class weight dihitung dari distribusi training. Checkpoint dengan validation AUC tertinggi digunakan untuk evaluasi data uji. Deep Ensemble dibentuk dari lima Light CNN yang dilatih secara independen.

F. Interpretabilitas dan Ketidakpastian

Grad-CAM diterapkan pada feature map konvolusi terakhir untuk memvisualisasikan area yang memengaruhi skor kelas [8]. Karena validitas penjelasan bergantung pada kemampuan prediksi, Grad-CAM pada penelitian ini hanya ditafsirkan sebagai diagnostik ketika macro AUC-ROC data uji berada pada atau di bawah referensi no-discrimination 0,50.

SHAP dihitung pada surrogate Random Forest yang menggunakan 40 fitur agregat, yaitu mean, standar deviasi, minimum, dan maksimum dari masing-masing 10 kanal. Surrogate ini bukan model CNN sehingga nilai SHAP hanya digunakan untuk melihat asosiasi global fitur agregat, bukan atribusi piksel langsung dan bukan bukti kausal [9].

Monte Carlo Dropout melakukan 50 stochastic forward passes dengan dropout tetap aktif untuk membentuk distribusi prediksi. Deep Ensemble menggabungkan lima Light CNN independen. Ketidakpastian diringkas menggunakan confidence, predictive entropy, normalized entropy, variation ratio, dan—untuk Monte Carlo Dropout—mutual information [9], [11].

$$H(p) = -\Sigma\, pk \log(pk); \quad Hnorm = H/\log(3)$$

G. Evaluasi dan Inferensi Spasial

Evaluasi klasifikasi yang terdokumentasi memakai accuracy, macro-F1, macro AUC-ROC dan confusion matrix pada sembilan sampel uji, tetapi interpretasinya bersyarat pada label acuan yang belum tervalidasi. Inferensi grid awal menggunakan 500 patch berurutan (step 128) dengan p01–p99 training-only dan softmax deterministik. Kode tersebut tidak menerapkan polygon land mask, tidak menolak patch pesisir berdasarkan footprint, dan mengganti NaN/Inf dengan nilai numerik setelah normalisasi; pengujian mean intensitas mendekati nol bukan pemeriksaan NoData yang memadai. Inferensi daratan final harus memakai polygon valid, memastikan seluruh footprint berada pada daratan, membuktikan validitas piksel tiap kanal sebelum normalisasi, dan mengambil sampel grid secara spasial representatif.

III. HASIL DAN PEMBAHASAN

A. Cakupan Data, Mosaik, dan Kontrol Leakage

Mosaik teknis 14 tile GeoTIFF menghasilkan raster 10 kanal dengan 54 lokasi yang memungkinkan pembacaan patch penuh. Namun, kelulusan pemeriksaan ukuran,

CRS, dtype, dan persentase nilai finite hanya menguji kesiapan numerik. Verifikasi daratan–laut dan keberadaan nilai NoData berkode angka belum tercakup pada audit semula. Oleh sebab itu, istilah "54 patch valid" pada hasil awal dibatasi pada validitas ekstraksi dan tidak digunakan sebagai bukti validitas geografis, fitur, atau label.

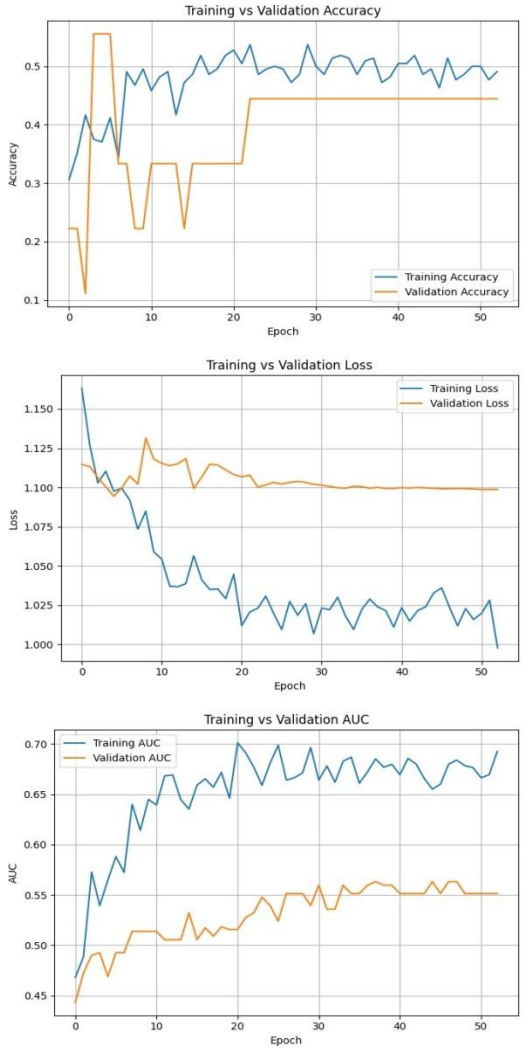
Audit lanjutan menargetkan tiga lapis bukti:

1. Asal-usul serta kriteria setiap label;
2. Kecocokan footprint dengan polygon daratan dan batas administratif kajian;
3. Mask NoData raster asli untuk semua kanal. Normalisasi p01–p99 berbasis training-only dan split tanpa overlap tetap digunakan, tetapi tidak menyelesaikan potensi kesalahan geografis maupun pelabelan. Rekonstruksi dataset, evaluasi Light CNN/ensemble, dan interpretasi XAI wajib diulang jika audit mengubah titik acuan atau piksel masukan.

B. Pelatihan Light CNN

Light CNN dilatih hingga 53 epoch sebelum early stopping. Validation AUC tertinggi sebesar 0,5632 dicapai pada epoch 38 dan disimpan sebagai checkpoint terbaik; pada epoch tersebut training accuracy 0,5139, training AUC 0,6852, validation accuracy 0,4444, dan validation loss 1,1000. Kurva pelatihan memperlihatkan peningkatan AUC training hingga kisaran 0,69, sementara validation AUC bertahan di sekitar 0,55. Pola ini menunjukkan pemisahan kelas pada data validasi masih terbatas meskipun data leakage telah dikoreksi.

Model CNN SE-Attention dari eksperimen awal tidak digunakan sebagai klaim hasil final karena pipeline revisi yang terdokumentasi melakukan pelatihan ulang pada Light CNN dan lima anggota Deep Ensemble berbasis arsitektur Light CNN. Dengan demikian, pembahasan kuantitatif setelah revisi difokuskan pada model yang benar-benar direrun menggunakan 54 patch, spatial split, dan scaler training-only.



Gambar 4. Kurva accuracy, loss, dan AUC Light CNN

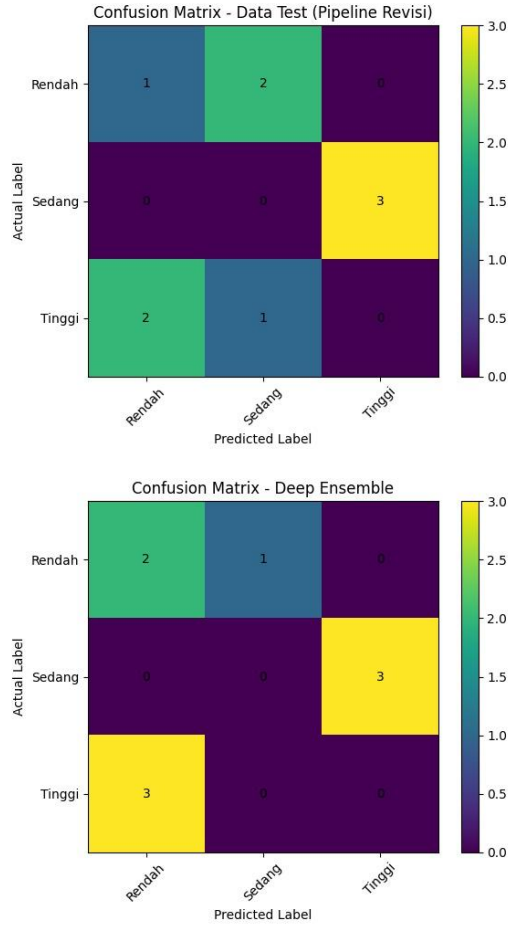
C. Evaluasi Klasifikasi pada Data Uji

Pada sembilan sampel uji yang seimbang (masing-masing tiga sampel per kelas), Light CNN hanya memprediksi satu sampel dengan benar. Accuracy, macro-precision, macro-recall, dan macro-F1 masing-masing sebesar 0,1111, sedangkan macro AUC-ROC one-vs-rest sebesar 0,3148. Nilai AUC tersebut berada di bawah referensi no-discrimination 0,50 dan, dengan ukuran uji yang hanya sembilan sampel, menunjukkan kemampuan diskriminasi yang lemah serta estimasi performa yang masih tidak stabil. Hasil ini tidak boleh dibaca sebagai bukti bahwa model secara universal lebih buruk dari tebakan acak, tetapi sebagai indikasi kuat

bahwa model belum memiliki pemisahan kelas yang memadai pada data uji spasial ini.

Tabel 4. Perbandingan performa klasifikasi pada sembilan sampel uji; MC Dropout menggunakan mean probability dari 50 stochastic forward passes.

Metrik	Light CNN	MC Dropout	Deep Ensemble
Accuracy	0,1111	0,1111	0,2222
Macro Precision	0,1111	0,1111	0,1333
Macro Recall	0,1111	0,1111	0,2222
Macro F1	0,1111	0,1111	0,1667
Macro AUC	0,3148	0,3704	0,4444



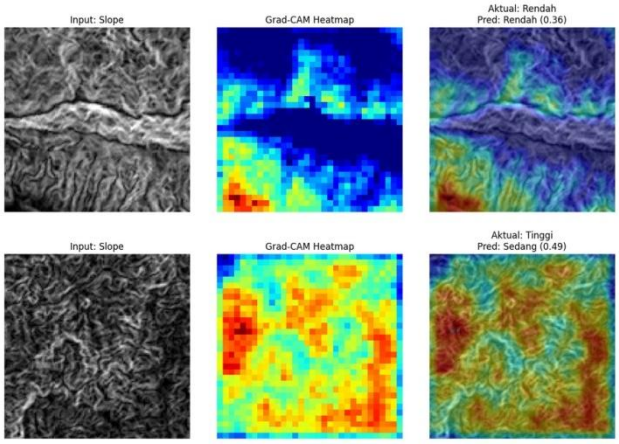
Gambar 5. Confusion matrix Light CNN (atas) dan Deep Ensemble (bawah) pada sembilan sampel uji pipeline

Prediksi rata-rata Monte Carlo Dropout juga hanya menghasilkan satu prediksi benar dari sembilan sampel; macro AUC-ROC berdasarkan mean probability sebesar 0,3704. Deep Ensemble memberikan dua prediksi benar dengan accuracy 0,2222, macro-F1 0,1667, dan AUC-ROC 0,4444. Walaupun sedikit lebih tinggi daripada Light CNN deterministik, ensemble tetap berada di bawah referensi 0,50. Penambahan stochastic inference maupun ensemble belum mengatasi keterbatasan representasi dan ukuran sampel.

D. Model melalui Grad-CAM dan SHAP

Grad-CAM dihitung menggunakan checkpoint Light CNN pipeline. Contoh yang ditampilkan mencakup satu prediksi benar kelas Rendah dan satu salah klasifikasi kelas Tinggi menjadi Sedang. Namun, karena macro AUC-ROC Light CNN hanya 0,3148, hotspot Grad-CAM diperlakukan sebagai visualisasi perilaku internal model. Pola perhatian tidak ditafsirkan sebagai lokasi penyebab atau bukti spasial faktor longsor.

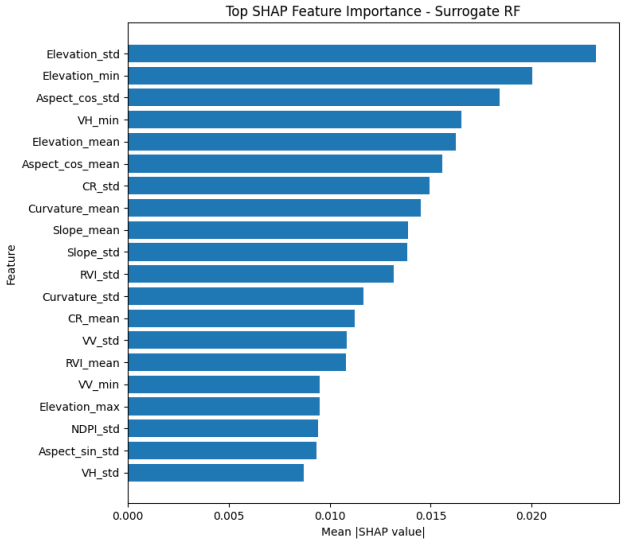
Pembatasan interpretasi tersebut secara langsung menjawab catatan reviewer: visualisasi XAI hanya bernilai substantif jika model terlebih dahulu menunjukkan kemampuan diskriminasi yang memadai. Pada kondisi saat ini, Grad-CAM berguna untuk mendiagnosis bagian patch yang digunakan model ketika membuat keputusan yang sebagian besar masih salah.



Gambar 6. Grad-CAM diagnostik pada pipeline, prediksi benar kelas Rendah (atas) dan salah klasifikasi Tinggi→Sedang (bawah).

Surrogate Random Forest menggunakan 40 fitur agregat dan menghasilkan accuracy 0,1111 pada data uji. Peringkat mean absolute SHAP tertinggi adalah Elevation_std (0,023166), Elevation_min (0,020045), Aspect_cos_std (0,018415), VH_min (0,016509), Elevation_mean (0,016223), Aspect_cos_mean (0,015560), CR_std (0,014952), Curvature_mean (0,014503), Slope_mean (0,013864), dan Slope_std (0,013836).

Kemunculan fitur topografi dan komponen Aspect_cos pada peringkat atas menunjukkan bahwa transformasi sirkular telah masuk ke analisis fitur. Namun, karena surrogate dan CNN sama-sama memiliki performa uji rendah, urutan SHAP ini hanya menunjukkan asosiasi internal pada dataset terbatas dan tidak dapat digunakan untuk menyimpulkan faktor kausal kerawanan longsor.



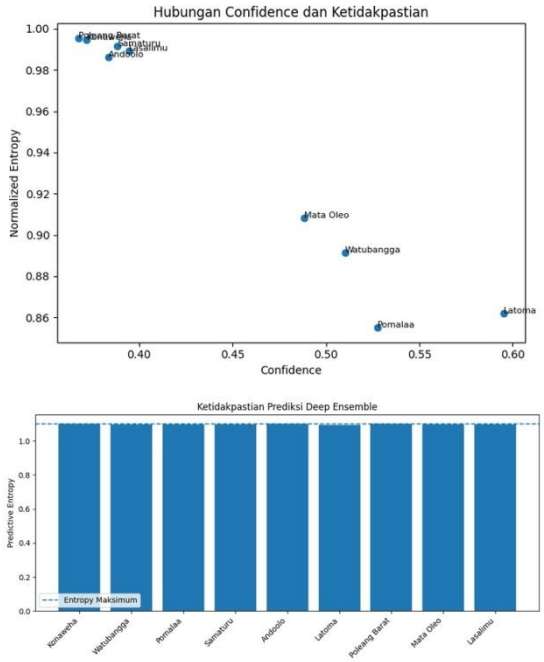
Gambar 7. Peringkat mean absolute SHAP pada surrogate Random Forest setelah penggunaan 10 kanal, termasuk Aspect_sin dan Aspect_cos.

E. Analisis Ketidakpastian Prediksi

Monte Carlo Dropout menghasilkan confidence rata-rata 0,4475, predictive entropy 1,0343, dan normalized entropy 0,9415. Rata-rata mutual information sebesar 0,0108 dan variation ratio 0,5525. Seluruh sembilan sampel dikategorikan sebagai ketidakpastian tinggi/perlu prioritas pemeriksaan. Deep Ensemble menghasilkan confidence rata-rata 0,3481, predictive entropy 1,0973, dan normalized entropy 0,9988, sangat dekat dengan entropy maksimum klasifikasi tiga kelas, log(3)=1,0986.

Tabel 5. Ringkasan ketidakpastian prediksi pada sembilan sampel uji setelah rerun pipeline.

Metode	Confidence	Predictive Entropy	Normalized Entropy
MC Dropout	0,4475	1,0343	0,9415
Deep Ensemble	0,3481	1,0973	0,9988



Gambar 8. Ketidakpastian pada pipeline, hubungan confidence–normalized entropy Monte Carlo Dropout (atas) dan predictive entropy Deep Ensemble (bawah).

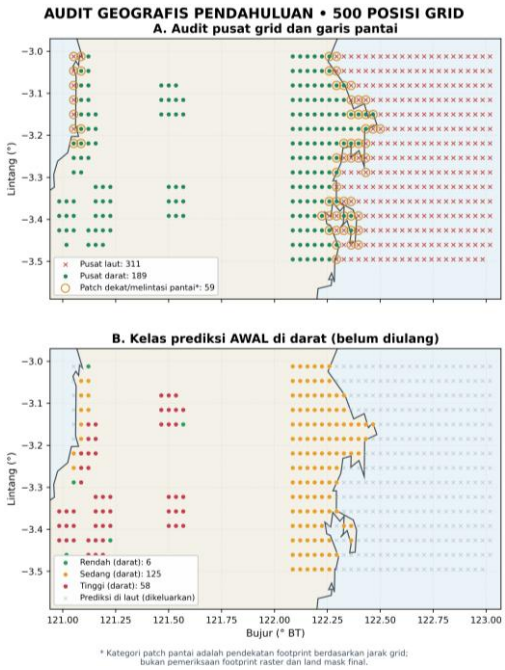
Nilai entropy yang tinggi konsisten dengan performa klasifikasi yang lemah. Dengan demikian, uncertainty tidak diposisikan sebagai penguat keputusan kelas, melainkan sebagai sinyal bahwa model belum yakin terhadap pemisahan tiga kelas. Dalam konteks mitigasi, keluaran tersebut lebih tepat digunakan untuk menandai lokasi yang memerlukan verifikasi data tambahan daripada sebagai dasar keputusan operasional.

F. Inferensi Spasial Berbasis Grid

Audit pendahuluan terhadap 500 posisi grid dari eksperimen lama dilakukan memakai polygon garis pantai GSHHS tingkat intermediate sebagai referensi. Sebanyak 189 pusat grid terindikasi di daratan (37,80%) dan 311 di perairan (62,20%). Dalam pemeriksaan tambahan dengan footprint yang didekati dari jarak antarpusat grid, 59 patch berpotensi memotong garis pantai (26 pusat pada daratan dan 33 di perairan). Kategori footprint ini dapat tumpang tindih dengan status pusat darat/laut dan bukan kategori ketiga yang dijumlahkan. Temuan tersebut bersifat awal karena batas patch belum direkonstruksi dari transformasi raster asli. Dari 311 pusat di perairan, seluruhnya diprediksi Sedang; pada 189 pusat yang terindikasi daratan, prediksi awal terdiri atas 6 Rendah, 125 Sedang, dan 58 Tinggi. Karena itu, proporsi kelas Sedang 87,20% pada Gambar 9 sebelumnya sangat dipengaruhi grid laut, urutan scanning, dan karakteristik prediksi model sehingga tidak dapat dianggap sebagai kerawanan daratan. Rata-rata normalized entropy untuk pusat di daratan dan perairan masing-masing sekitar 0,9449 dan 0,9292; tingginya kedua nilai menunjukkan ketidakpastian prediksi belum teratasi.

Tabel 6. Distribusi kelas pada 500 pusat patch grid eksperimental setelah pipeline.

Kelas	Jumlah	Persentase
Rendah	6	1,20%
Sedang	436	87,20%
Tinggi	58	11,60%



Gambar 9. Geografis PENDAHULUAN 500 Pusat Grid [12].

Tahap mensyaratkan polygon daratan beresolusi memadai dan metadata sumber, pemeriksaan titik pusat dan footprint patch pada sistem koordinat yang sama, validasi mask NoData dan nilai non-finite pada 10 kanal sebelum normalisasi, lalu inferensi ulang pada patch yang sepenuhnya berada di daratan.

Seleksi pusat grid harus mencakup wilayah studi secara representatif, bukan 500 posisi valid pertama menurut urutan raster. Hasil akhir wajib menampilkan peta daratan dan garis pantai, jumlah titik darat/laut/pesisir, distribusi kelas, confidence, entropy, serta berkas audit penolakan patch. Hingga audit raster dan rerun tersebut selesai, Gambar 9 hanya dapat digunakan sebagai bukti diagnostik kesalahan pemilihan grid, bukan representasi kerawanan Sulawesi Tenggara.

G. Implikasi dan Keterbatasan

Kontribusi penelitian berupa mosaik raster, normalisasi training-only, spatial split tanpa overlap, serta evaluasi XAI/ketidakpastian tetap memiliki nilai metodologis. Namun, audit geospasial mengidentifikasi 311/500 pusat prediksi yang terindikasi di laut dan potensi 59 patch

pesisir. Dominasi kelas Sedang pada 500 grid lama terutama tidak dapat dipisahkan dari masalah pemilihan lokasi dan kemampuan diskriminasi Light CNN yang rendah. Capaian awal harus dipisahkan secara tegas dari keluaran final hasil inferensi ulang yang belum tersedia.

Validitas label juga merupakan keterbatasan fundamental: daftar 54 koordinat/kelas diketik dalam sel notebook, sementara dokumen penyedia, waktu perolehan, jenis titik, dasar pengkelasan, dan validasi independen belum terdokumentasi. Sekurangnya satu koordinat inventaris terindikasi membutuhkan pemeriksaan garis pantai. Tindakan korektif yang wajib dilakukan adalah memperoleh data primer dari instansi penyedia, mencatat metadata per fitur, memvalidasi koordinat dan label, lalu menjalankan kembali split, fitting scaler, pelatihan, evaluasi, serta inferensi land-only. Tanpa tahap ini, interpretasi model dan peta tetap bersifat eksploratif.

IV. KESIMPULAN DAN SARAN

Penelitian ini berhasil mengembangkan kerangka pemrosesan dan evaluasi kerawanan longsor berbasis Explainable Deep Learning melalui integrasi citra Sentinel-1 GRD dan turunan topografi dalam 10 kanal masukan. Sebanyak 54 patch berukuran 128 × 128 piksel berhasil diekstraksi dari mosaik raster dan dibagi menggunakan pendekatan berbasis blok spasial menjadi 36 data latih, 9 data validasi, dan 9 data uji. Penerapan normalisasi robust scaling p01–p99 yang dihitung secara eksklusif dari data latih menjadi kontribusi metodologis dalam mengurangi risiko kebocoran informasi serta meningkatkan konsistensi, keterlacakan, dan reproduksibilitas eksperimen.

Hasil evaluasi menunjukkan bahwa Light CNN memperoleh accuracy sebesar 0,1111 dan macro AUC-ROC sebesar 0,3148 pada sembilan data uji spasial, yang mengindikasikan kemampuan diskriminasi model masih terbatas. Analisis geografis terhadap 500 pusat grid menggunakan garis pantai GSHHS resolusi menengah mengidentifikasi 189 titik yang terindikasi berada di daratan dan 311 titik di perairan, sedangkan sekitar 59 patch diperkirakan beririsan dengan garis pantai. Temuan ini menunjukkan pentingnya kesesuaian lokasi pengamatan dengan karakteristik geografis wilayah kajian untuk meningkatkan validitas interpretasi hasil pemetaan.

Kontribusi penelitian juga mencakup penerapan Grad-CAM dan SHAP untuk menganalisis perilaku prediksi model, serta Monte Carlo Dropout dan Deep Ensemble untuk mengestimasi ketidakpastian klasifikasi. Integrasi pendekatan tersebut memberikan kerangka evaluasi yang tidak hanya mempertimbangkan keluaran kelas prediksi, tetapi juga interpretabilitas dan tingkat keyakinan model. Meskipun demikian, hasil interpretasi masih bersifat diagnostik karena keterbatasan kemampuan model dan perlunya verifikasi lebih lanjut terhadap sumber serta kriteria pelabelan 54 titik acuan.

Secara keseluruhan, kontribusi utama penelitian adalah pengembangan kerangka analisis kerawanan longsor berbasis Explainable Deep Learning yang mengintegrasikan pengendalian kebocoran informasi, evaluasi spasial, interpretabilitas model, dan pengukuran ketidakpastian prediksi secara sistematis, transparan, serta dapat direproduksi. Temuan penelitian memberikan landasan metodologis bagi pengembangan sistem pemetaan kerawanan longsor yang lebih terukur dan dapat dipertanggungjawabkan secara ilmiah. Pengembangan selanjutnya diarahkan pada penguatan validitas data acuan, penyempurnaan cakupan geografis wilayah analisis, peningkatan representativitas sampel, serta evaluasi kemampuan generalisasi dan kalibrasi ketidakpastian model. Dengan demikian, penelitian ini memberikan kontribusi bagi pengembangan metode analisis geospasial berbasis kecerdasan buatan untuk mendukung kajian mitigasi bencana tanah longsor di Sulawesi Tenggara.

DAFTAR PUSTAKA

[1] F. Guzzetti, A. C. Mondini, M. Cardinali, F. Fiorucci, M. Santangelo, and K.-T. Chang, "Landslide inventory maps: New tools for an old problem," *Earth-Science Rev.*, vol. 112, no. 1–2, pp. 42–66, 2012.

[2] C. Cakra, B. Baharuddin, A. M. Islah, and L. O. M. B. Aksara, "Klasifikasi Daerah Rawan Longsor menggunakan Metode Deep Learning Berbasis Data Citra Sentinel-1," *Semant. Tek. Inf.*, vol. 11, no. 2, 2025.

[3] M. Occhipinti, F. Carboni, S. Amorini, N. Paltriccia, C. López-Martínez, and M. Porreca, "Implementing the European space agency's sentinel application platform's open-source python module for differential synthetic aperture radar interferometry coseismic ground deformation from Sentinel-1 data," *Remote Sens.*, vol. 16, no. 1, p. 48, 2023.

[4] F. F. Rahman, "Identifikasi Longsor dengan Menggunakan Citra Sentinel 1 SAR Pada Sub DAS Mamasa, DAS Saddang= Landslide Identification Using Sentinel 1 SAR Imagery in Mamasa Sub-Watershed, Saddang Watershed," 2022, *Universitas Hasanuddin*.

[5] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[6] E. P. Umar *et al.*, "Kajian Mitigasi Bencana Tanah Longsor Ruas Jalan Meluhu-Lasolo, Sulawesi Tenggara," *J. Geocelebes*, vol. 3, no. 2, pp. 51–58, 2019.

[7] B. N. P. Bencana, "Metodologi InaRISK," Nov. 02, 2026, *Badan Nasional Penanggulangan Bencana (BNPB)*. [Online]. Available: https://inarisk.bnpb.go.id/v4/metodologi

[8] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.

[9] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.

[10] Y. Gal and Z. Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," in *international conference on machine learning*, PMLR, 2016, pp. 1050–1059.

[11] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.

[12] P. Wessel and W. H. F. Smith, "A global, self-consistent, hierarchical, high-resolution shoreline database," *J. Geophys. Res. Solid Earth*, vol. 101, no. B4, pp. 8741–8743, 1996.