



RemoteSense: Burn-scar-based crop residue burning detection and alerting with a fine-tuned vision language model on Sentinel-2 imagery

Sahil Patil^a, Amol Powar^a, Sarvesh Parab^a, Prof. Revti Jadhav^b

^aComputer Engineering, Smt. Indira Gandhi College of Engineering, Ghansoli, Maharashtra, India

E-mail address: 2023ce41f@sigce.edu, 2024ce138d@sigce.edu.in, 2023ce38f@sigce.edu.in

ABSTRACT:

Crop residue burning in northern India causes recurring air-quality crises, yet active-fire sensors (MODIS, VIIRS) miss most events: their fixed overpass times are increasingly avoided by farmers, and a CEEW study found VIIRS detected only 7 of 169 burnt fields (4.1%) in Sangrur over two days. Because burn scars persist long after flames die, we propose detecting scar evidence rather than the fire itself. We present RemoteSense, an end-to-end platform that (i) fine-tunes a 450M-parameter vision language model (LFM2.5-VL-450M) with LoRA to read paired RGB and SWIR Sentinel-2 composites and return a structured JSON burn assessment; (ii) corroborates detections with live NASA FIRMS hotspots; (iii) dispatches severity-routed, geo-tagged alerts via SMS, email, web push, and webhooks; and (iv) exports the model to quantized ONNX for browser and edge inference. On 178 clear-sky, district-disjoint test tiles, fine-tuning raises mean five-field accuracy from 21.9% to 74.8% and severity accuracy from 11.8% to 53.9%; binary burn detection reaches 73.6% versus a 65.2% majority-class baseline. These results are preliminary: labels are semi-automatically generated, the test set is single-season, and no comparison with spectral-index or CNN baselines is yet available.

Keywords: vision language model, crop residue burning, Sentinel-2, burn scar detection, LoRA fine-tuning, edge AI, alert system

1. Introduction

Open burning of paddy residue across the Indo-Gangetic Plain is a recurring source of particulate pollution during October–November, and smoke from Punjab and Haryana is carried toward densely populated areas such as Delhi (Anand et al., 2025). Enforcement depends on knowing which fields were burnt, and that is where current monitoring falls short.

The timing gap. Official fire counts rely on active-fire detections from MODIS and VIIRS. VIIRS on Suomi NPP and NOAA-20 observes the Indo-Gangetic Plain at roughly 01:30 and 13:30 local time, and MODIS on Terra and Aqua at about 10:30 and 13:30 (CEEW, 2025). A fire that starts in the late afternoon and dies out by evening can fall entirely between overpasses. Analyses of 15-minute geostationary data report that over 90% of large farm fires in Punjab in 2024 and 2025 occurred after 3 PM, compared with 3% in 2021 (The Tribune, 2025), and an ISRO study reached a similar conclusion for 2020–2024 (ThePrint, 2025). CEEW quantified the consequence on a small sample: in a two-day, 74 km² analysis of very-high-resolution imagery over Sangrur (November 2020), 169 fields were fully or partially burnt, but VIIRS detected only 7 fires and MODIS none, i.e., a recall of 7/169 = 4.1% (CEEW, 2025). This is a case study rather than a national estimate, but it shows how large the undercount can be.

Key idea. Flames last minutes to hours, whereas burn scars remain visible for days until tillage or sowing. Sentinel-2 provides 10 m multispectral imagery with a roughly five-day revisit from a constellation (Drusch et al., 2012), so a scar can be observed regardless of when the fire was lit. Scar mapping is not trivial: in smallholder landscapes, burn scars are easily confused with bare or ploughed soil, and spectral-index thresholds produce many false positives (Anand et al., 2025; Deshpande et al., 2022). Deep-learning segmentation has improved on index thresholds (Anand et al., 2025; Knopp et al., 2020), but these models output pixel masks. They do not report the field-level context that an enforcement officer needs: how severe the burn looks, whether smoke is visible, what stage of the crop cycle the field is in, and whether the image is good enough to trust.

Approach. Vision language models (VLMs) can produce structured, contextual descriptions from imagery (Li et al., 2024; Soni et al., 2025). We fine-tune a small VLM to read paired RGB and SWIR composites and emit a fixed JSON schema, and wrap it in a platform that corroborates detections with NASA FIRMS and routes alerts to enforcement officers. To the best of our knowledge, no prior work applies a fine-tuned VLM to structured assessment of crop-residue burn scars from Sentinel-2.

Contributions. This paper makes the following contributions:

- A LoRA-fine-tuned LFM2.5-VL-450M that maps RGB+SWIR Sentinel-2 composites to structured burn assessments, trained on 1,103 labelled image pairs and evaluated on district-disjoint tiles.
- A multi-source pipeline combining scar-based VLM assessment with live NASA FIRMS hotspots for corroboration.
- A severity-routed alert engine (SMS, email, web push, webhook) with audit logging and escalation.

- A quantized ONNX export that runs in the browser (WebGPU) and is a candidate for on-board inference, together with an honest account of what has and has not been validated.

The rest of the paper is organised as follows. Section 2 reviews the literature. Section 3 describes how the review of literature was conducted and the RemoteSense methodology. Section 4 presents the findings of the review, the research gaps and the resulting choice of approach, and the experimental results. Section 5 concludes.

2. Literature review

2.1. Active-fire monitoring

VIIRS 375 m active-fire data (Schroeder et al., 2014) and MODIS 1 km products underpin India's official counts, but they only capture fires burning during an overpass, and coarse pixels miss small fires (CEEW, 2025; Ramo et al., 2021).

2.2. Burn-scar mapping

The Normalized Burn Ratio, $(NIR - SWIR2)/(NIR + SWIR2)$ (Key & Benson, 2006), its differenced form (Miller & Thode, 2007), MIRBI (Trigg & Flasse, 2001) and the Sentinel-2-specific BAIS2 (Filipponi, 2018) are standard. For Punjab, Walker et al. (2022) combined PlanetScope and Sentinel-2 with a random forest and reported 82% accuracy against ground-based monitoring; Deshpande et al. (2022) built an index-based smallholder burnt-area method for Sentinel-2; and Anand et al. (2025) fine-tuned a U-Net on hand-annotated Punjab data, reaching IoU 0.54 and Dice 0.64 while detecting patches under 250 m². U-Net variants for Sentinel-2 burned-area segmentation are also established (Knopp et al., 2020). These methods produce masks or binary labels; none returns contextual attributes such as smoke, crop phase or image-quality self-assessment.

2.3. Vision language models for Earth observation and fire

GeoChat (Kuckreja et al., 2024), EarthDial (Soni et al., 2025), which handles multispectral Sentinel-2 input, and GeoLLaVA-8K (Wang et al., 2025), which addresses token explosion for ultra-high-resolution scenes with background-token pruning and anchored token selection, show that VLMs can be adapted to remote sensing. In the fire domain, WildfireVLM pairs a YOLOv12 detector on Landsat/GOES imagery with multimodal LLMs that generate risk assessments (Ayanzadeh et al., 2026), and ForestFireVLM fine-tunes a 7B VLM to describe forest-fire scenes in UAV images (Seidel et al., 2025). We are not aware of work that targets crop-residue burn scars with a fine-tuned VLM.

2.4. Parameter-efficient fine-tuning and on-board AI

LoRA (Hu et al., 2022) and QLoRA (Dettmers et al., 2023) adapt large models by training low-rank updates. On the platform side, ESA's Φsat-2, launched in August 2024, runs AI apps (including cloud and wildfire detection) on an Intel Myriad 2 processor (European Space Agency, n.d.-a), and TakeMe2Space's MOI-TD demonstrated on-orbit AI inference in 2024–25, with the Orin NX-based MOI-1A having been scheduled for launch on 1 October 2026 (Reuters, 2026).

2.5. Comparison of approaches

Table 1 summarizes how the reviewed approaches differ.

Table 1 - Qualitative comparison of approaches (from published capabilities).

Approach	Post-fire detection	Field context	Structured output	Alerting
VIIRS/MODIS active fire	No	No	No	Counts only
Geostationary (SEVIRI)	Partly	No	No	No
Spectral indices	Yes	No	No	No
DL segmentation (Anand et al., 2025; Knopp et al., 2020)	Yes	Limited	Mask	No
Generic RS-VLM (Soni et al., 2025)	Yes	Yes	Partial	No
RemoteSense (this work)	Yes	Yes	Yes	Yes

3. Methodology

3.1. Conduct of the review of literature

The review of literature was carried out to establish (a) how well existing monitoring detects crop residue burning, (b) what burn-scar mapping methods achieve in smallholder landscapes, (c) whether vision language models have been applied to burn-scar or fire assessment, and (d) which adaptation and

deployment techniques make a small model practical for operational use. Sources were grouped under these four questions plus a fifth group of data and platform documentation (Table 2).

The corpus comprises 28 sources published between 2001 and 2026: 16 peer-reviewed journal, conference and proceedings papers, two preprints, one technical report, one institutional report (CEEW), three news reports used for recent developments not yet covered by the journal literature, and five official data, platform or model-card documents. Studies were retained when they reported quantitative results or capabilities for burned-area or active-fire detection, Sentinel-2 or comparable multispectral imagery, vision language models for Earth observation or fire, or parameter-efficient and on-board inference. Findings drawn from news and institutional reports are presented as case studies or reported findings, not as peer-reviewed evidence. Published performance figures were compared qualitatively only, because the studies use different data, tasks and metrics.

Each study was then examined for what it outputs (mask, label, count or text), whether it provides field-level context, and whether it supports alerting. The resulting research gaps were mapped to design decisions for RemoteSense (Section 4.2).

Table 2 - Organisation of the reviewed literature by theme.

Theme	Focus	Key sources
Monitoring gap	Overpass timing; undercount of crop fires	CEEW (2025); The Tribune (2025); ThePrint (2025); Schroeder et al. (2014); Ramo et al. (2021)
Burn-scar indices and segmentation	Spectral indices; smallholder and Punjab mapping	Key & Benson (2006); Trigg & Flasse (2001); Miller & Thode (2007); Filipponi (2018); Knopp et al. (2020); Deshpande et al. (2022); Walker et al. (2022); Anand et al. (2025)
VLMs for Earth observation and fire	Remote-sensing VLMs; wildfire VLMs	Li et al. (2024); Kuckreja et al. (2024); Soni et al. (2025); Wang et al. (2025); Seidel et al. (2025); Ayanzadeh et al. (2026)
Efficient adaptation and on-board AI	LoRA and QLoRA; on-orbit inference; compact VLM	Hu et al. (2022); Dettmers et al. (2023); European Space Agency (n.d.-a); Reuters (2026); Liquid AI (2026)
Data and platforms	Sentinel-2, CDSE, FIRMS, SimSat	Drusch et al. (2012); European Space Agency (n.d.-b); NASA (n.d.); DPhi Tech (n.d.)

3.2. System overview and imagery pipeline

RemoteSense has four layers: data acquisition, VLM inference, alerting, and a web interface (Fig. 1). For a query point, the backend computes a 5 km × 5 km bounding box (longitude span corrected by $\cos \phi$), authenticates to the Copernicus Data Space Ecosystem (European Space Agency, n.d.-b) via OAuth2 with token caching, and queries the STAC catalogue for the most recent Sentinel-2 L2A scene with less than 60% cloud cover within a configurable window (default 60 days). Three 256 × 256 composites are rendered server-side via evalscripts. Since 5 km at native 10 m resolution corresponds to 500 px, the 256-pixel tiles are resampled to approximately 19.5 m per pixel.

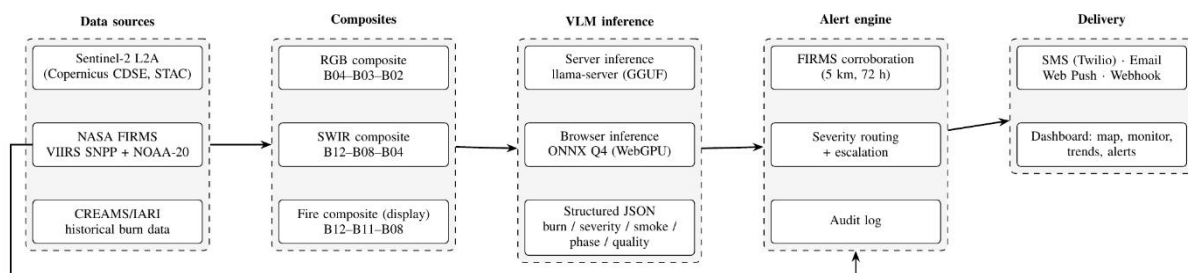


Fig. 1 - RemoteSense architecture. Sentinel-2 imagery is converted to RGB and SWIR composites, assessed by the VLM (server or browser), corroborated with FIRMS hotspots, routed by severity and delivered over four channels while every step is logged.

- RGB (B04-B03-B02) with a 3.5× gain for visibility.
- SWIR (B12-B08-B04), a false-colour composite in which healthy crop is bright green, stubble and bare soil are brown, and burn scars are charcoal to dark red with straight field edges. Pixels with $NBR < -0.10$ are progressively darkened to accentuate scars.
- Fire (B12-B11-B08), highlighting pixels with $B12 > 0.35$, $B12 > 1.2 \times B11$ and $B12-B08 > 0.15$. These thresholds were tuned empirically and are used for map display, not as VLM input.

The returned scene date is always shown to the user. Since the scene can be days or weeks old, the pipeline is near-real-time with respect to imagery availability, not with respect to the fire itself, and scars may already be obscured by tillage or sowing (Anand et al., 2025).

3.3. Live hotspot layer

NASA FIRMS provides near-real-time VIIRS 375 m detections (NASA, n.d.; Schroeder et al., 2014). The platform fetches CSV data for India from both SNPP and NOAA-20, discards low-confidence detections, and aggregates counts and fire radiative power by state over 1–5 day windows. FIRMS

serves two roles: a live monitoring map, and a corroboration signal for the alert engine.

3.4. VLM inference and output schema

Both composites are passed to the model with a system prompt that sets a Sentinel-2 analyst role and a user prompt that states each image's band combination, gives the SWIR colour-to-feature guide above, and fixes the JSON schema. Two pathways share the prompt: (i) a GPU server running llama-server (OpenAI-compatible API, temperature 0, 512 max tokens), and (ii) in-browser inference with Transformers.js on WebGPU using the Q4-quantized ONNX export, which allows server-free operation. A tolerant JSON extractor (first { to last }, placeholder normalization) parses outputs into a typed object. The supervised labels and all reported evaluation cover five fields (Table 3); the deployed prompt additionally requests burn_fraction, burn_freshness and fire_detected, which were not part of fine-tuning or evaluation and are treated as advisory only.

Table 3 - Output fields used in training and evaluation.

Field	Type	Values
burn_detected	bool	scar visible in tile
burn_severity	enum	none / low / moderate / high
active_smoke_visible	bool	plume visible
vegetation_phase	enum	pre-harvest / harvesting / post-harvest / fallow
image_quality_limited	bool	cloud or missing data

3.5. Alert engine

When burn_detected is true, the engine (Fig. 2) (1) queries FIRMS hotspots within 5 km over the preceding 72 h and labels the alert corroborated or uncorroborated; (2) routes by burn_severity; (3) dispatches over the configured channels; and (4) logs delivery, acknowledgment and field-verification outcome. Defaults, all configurable per district, are: high severity to district enforcement officers and pollution-control-board teams by SMS and push; moderate to block-level officers by email and push; low to the dashboard and a periodic email digest; per-zone rate limiting; and escalation of unacknowledged high-severity alerts after two hours. SMS (Twilio) carries coordinates, severity and a dashboard link; email includes both composites and the JSON assessment; push uses Web Push with VAPID; webhooks forward the JSON payload to external systems.

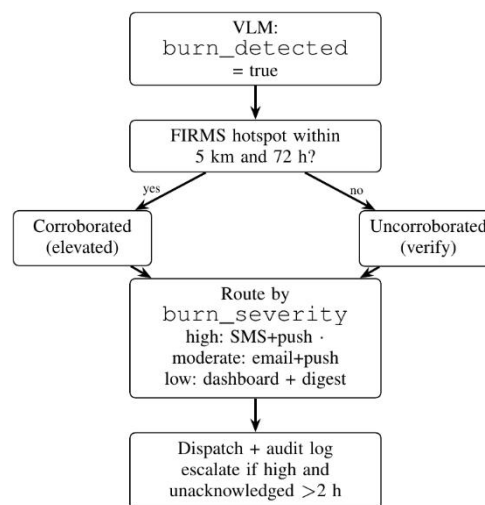


Fig. 2 - Alert decision flow.

Corroboration has an inherent limitation: a fire lit after the last overpass, which is exactly the case motivating this work, cannot be corroborated by FIRMS. Uncorroborated alerts are therefore still sent, but flagged pending field verification. No alert is intended to support a penalty without physical verification.

3.6. Web platform

The interface (Next.js/React, MapLibre GL, Recharts) offers a click-to-analyse map, a live FIRMS monitor, 2020–2025 district-level trend charts from CREAMS/IARI data, and an alert dashboard for dispatch logs, recipients, zones and acknowledgments. District trend views currently cover 23 tracked districts.

3.7. Dataset

Tiles were sampled over 55 districts in five states (Punjab, Haryana, Uttar Pradesh, Madhya Pradesh, Rajasthan), from both fire-prone (e.g., Sangrur, Patiala, Ludhiana) and fire-free regions, using Sentinel-2 L2A data retrieved through the DPhi SimSat platform (DPhi Tech, n.d.) for March–November 2025. Each sample is a co-registered RGB+SWIR pair with a five-field JSON label (Table 3). Labels were produced in three stages: (1) automatic pre-labelling from NBR/BAIS2/dNBR thresholds and FIRMS hotspot cross-referencing (± 72 h); (2) active-fire pixel counting from fire composites; and (3) visual verification by annotators who resolved ambiguous cases (e.g., dark bare soil versus burn scar). Labels are expert-verified but not field-verified ground truth. Table 4 gives dataset statistics.

Table 4 - Dataset statistics.

Item	Value
Samples	1,370 pairs: 1,103 train / 267 test
Split	Geographic; no district shared across splits
Evaluated subset	178 clear-sky tiles from the test split
Classes (train)	718 no-burn / 385 burn (34.9% burn)
Severity (all)	none 65%, low 12%, moderate 15%, high 8%
Test severity counts	none 116, low 21, moderate 27, high 14
Tile	5 km $\times$ 5 km, 256 $\times$ 256 px, 2 composites

3.8. Model and fine-tuning

We use LFM2.5-VL-450M (Liquid AI, 2026), a 450M-parameter VLM that pairs an 86M-parameter SigLIP2 vision encoder with a 350M-parameter language model whose hybrid backbone interleaves short-convolution and grouped-query-attention layers. We chose it for its size, native multi-image input, reliable JSON generation and edge suitability. We adapt it with LoRA (Hu et al., 2022) (rank 32, $\alpha = 32$, dropout 0.05) on attention projections, feed-forward layers and the multimodal projector, giving 16.3M trainable parameters out of 465M (3.5%). Training used 5 epochs, learning rate 2×10^{-5} with cosine decay and 5% warm-up, effective batch size 16 (micro-batch 1, 16 accumulation steps), BF16, on a single 16 GB RTX A4000 in about 16 minutes. The loss is class-weighted: the no-burn class receives weight $718/385 \approx 1.86$, which penalizes false alarms more than misses, a deliberate choice given the cost of falsely accusing a farmer.

3.9. Evaluation protocol

We evaluate the base and fine-tuned models on 178 clear-sky tiles from districts unseen in training, scoring each field by exact-match accuracy and the overall score as the mean over the five fields. Both models use the same prompt and greedy decoding. We report a single training run; no confidence intervals across seeds are available, so we give Wilson 95% intervals for the binomial uncertainty of the 178-tile test set.

4. Results and discussion

4.1. Findings of the review of literature

Four findings emerge from the reviewed literature.

- **Active-fire products undercount crop burning.** Late-afternoon burning falls between VIIRS and MODIS overpasses; the CEEW case study recovered only 7 of 169 burnt fields (4.1%), and geostationary analyses place over 90% of large Punjab fires after 3 PM in 2024–2025.
- **Scar-based methods work but are limited in output.** Index thresholds confuse scars with bare or ploughed soil; the best Punjab-specific deep-learning result reviewed reaches IoU 0.54 and Dice 0.64 (Anand et al., 2025) and a random forest reaches 82% accuracy (Walker et al., 2022), but all return masks or binary labels without field-level context. These figures use different data, tasks and metrics and are not directly comparable.
- **VLMs are established in remote sensing but not for crop-residue scars.** Remote-sensing VLMs (Kuckreja et al., 2024; Soni et al., 2025; Wang et al., 2025) and fire-domain VLMs (Ayanzadeh et al., 2026; Seidel et al., 2025) exist, yet none targets structured assessment of crop-residue burn scars.
- **Compact adaptation and on-board inference are feasible.** LoRA and QLoRA make fine-tuning practical on a single GPU, and flight demonstrations of on-board AI (European Space Agency, n.d.-a; Reuters, 2026) make edge inference a realistic target.

4.2. Milestone and outcome of the review: research gaps and selection of approach

The review leads to six research gaps. Each gap determined a specific choice in the RemoteSense design (Table 5). Taken together, the outcome of the review is the selection of a small, LoRA-fine-tuned vision language model that reads Sentinel-2 RGB and SWIR composites and emits a structured JSON assessment, embedded in a platform that corroborates detections with FIRMS and routes alerts by severity.

Table 5 - Research gaps identified in the review and the approach selected.

Research gap	Evidence from the review	Approach selected for this work
G1. Fires missed by overpass-based sensors	VIIRS 7/169 (4.1%) in the CEEW case; most large fires after 3 PM (CEEW, 2025; The Tribune, 2025)	Detect burn-scar evidence in Sentinel-2 imagery rather than the fire itself; scars persist for days and the revisit is about five days
G2. Index thresholds confuse scars with bare soil	Many false positives in smallholder landscapes (Anand et al., 2025; Deshpande et al., 2022)	Paired RGB and SWIR (B12-B08-B04) composites with NBR-based darkening, read by a learned model; labels visually verified by annotators
G3. Outputs are masks, not field-level context	Segmentation and index methods give masks or binary labels only (Knopp et al., 2020; Anand et al., 2025)	VLM that returns a fixed JSON schema: burn, severity, smoke, vegetation phase and image-quality flag
G4. No fine-tuned VLM for crop-residue scars; large models are hard to deploy	RS-VLMs are general-purpose or forest-fire specific (Soni et al., 2025; Seidel et al., 2025)	LFM2.5-VL-450M chosen for size, multi-image input, reliable JSON and edge suitability; LoRA (rank 32, 3.5% of parameters); class-weighted loss to penalise false alarms
G5. Detections are not tied to an enforcement workflow	Active-fire products provide counts only (Table 1)	FIRMS corroboration plus severity-routed alerts over SMS, email, push and webhook, with audit log; uncorroborated alerts flagged for field verification
G6. Server dependence and raw-data downlink	On-board AI demonstrated in orbit (European Space Agency, n.d.-a; Reuters, 2026)	Q4-quantized ONNX export for browser (WebGPU) and edge inference, about 450 MB versus about 1.8 GB at FP32

One gap remains open after this work: the VLM has not yet been compared with an NBR/BAIS2 threshold classifier or a CNN trained on the same tiles (Section 4.3).

4.3. Experimental results

Table 6 reports accuracy per field. The image-quality column is derived from the reported overall score, which equals the mean of the five fields.

Table 6 - Accuracy (%) on 178 held-out clear-sky tiles.

Model	Burn	Severity	Smoke	Veg. phase	Quality	Overall
Base	73.6	11.8	0.0	24.2	0.0	21.9
Fine-tuned	73.6	53.9	100.0	46.6	100.0	74.8
Δ (pp)	0.0	+42.1	+100.0	+22.4	+100.0	+52.9
95% CI (FT)	66.7–79.5	46.6–61.1	–	39.4–54.0	–	–

The mean of burn detection and severity, the two fields most relevant to enforcement, rises from 42.7% to 63.8%. We draw four observations, with caveats.

Burn detection is unchanged by fine-tuning. Both models score 73.6% (131/178). A trivial classifier that always predicts “no burn” scores 65.2% (116/178 tiles carry severity “none”), so the improvement over the majority baseline is only 8.4 pp and confidence intervals overlap. We therefore withhold claims of reliable binary burn detection pending further evaluation.

Severity accuracy improves substantially over the base model but remains modest. Base severity accuracy is 11.8%, and fine-tuning raises it to 53.9%. This is still below the 65.2% of always predicting “none”; because that constant predictor is uninformative about burn presence, overall accuracy is a poor summary statistic on this imbalanced dataset, and macro-averaged metrics and a full confusion matrix are required (Section 4.6). With only 14 high-severity and 21 low-severity test tiles, per-class estimates carry high variance.

The 100% smoke and quality scores should not be over-interpreted. The test subset is clear-sky by construction, so positive examples of image_quality_limited and, we expect, of active_smoke_visible have little or no support. These scores indicate mainly that fine-tuning removed the base model’s systematic false-positive outputs for these fields, not that it recognizes smoke reliably.

Vegetation phase is the weakest field (46.6%). Distinguishing post-harvest from fallow, or pre-harvest from harvesting, needs temporal context that a single-date pair lacks.

Missing baselines. We have not compared against an NBR/BAIS2 threshold classifier or a CNN trained on the same tiles. Without them, we cannot claim that the VLM outperforms conventional scar detectors. Published numbers for other methods (Anand et al., 2025; Walker et al., 2022) use different data, tasks and metrics and are not comparable to ours.

4.4. Inference latency

Table 7 lists single-tile latency, including pre-processing, generation and JSON parsing. The Jetson figure is a projection, not a measurement.

Table 7 - Single-tile inference latency.

Deployment	Hardware	Latency
llama-server	RTX A4000	2–4 s
ONNX, WebGPU (browser)	consumer GPU	5–10 s
ONNX, CPU (browser)	modern x86 CPU	15–30 s
ONNX, edge (projected)	Jetson Orin NX	~1 s

4.5. Edge deployment concept

The fine-tuned model is exported to ONNX with 4-bit weight quantization for the vision encoder and decoder and FP32 embeddings, giving about 450 MB versus about 1.8 GB at FP32. This runs in WebGPU browsers and on ONNX Runtime CPU/CUDA providers, and is a candidate for an Orin NX-class module, which several commercial platforms now fly (Reuters, 2026). On-orbit operation has not been demonstrated; the following is a concept analysis.

A satellite running the model would downlink a JSON alert of about 200 bytes per detection and nothing for non-burn tiles. For comparison, a $5 \text{ km} \times 5 \text{ km}$ tile with all 13 MSI bands at native resolution and 16-bit depth is about 2.8 MB (computed from the band resolutions), a reduction of roughly 10^4 per alert. Over the Indo-Gangetic Plain (about $5 \times 10^5 \text{ km}^2$, or ~20,000 tiles per full coverage), and assuming a 2% burn rate, this is about 400 alerts (~80 KB) rather than ~56 GB of raw tiles. These figures illustrate the order of magnitude and depend on the assumed burn rate and on the model's false-alarm rate, which we have not yet measured separately.

4.6. Limitations

- **Label provenance.** Labels derive from spectral-index pre-labelling, FIRMS cross-referencing and visual review, with no field verification. The model may partly learn the biases of the labelling pipeline, and the 65/35 class balance reflects our sampling, not the true prevalence. Field-verified ground truth, as used by Anand et al. (2025) and Walker et al. (2022), is needed.
- **Scale and scope.** Training used 1,103 pairs and evaluation 178 tiles from one season (2025). District-disjoint splitting does not test temporal generalization.
- **Evaluation.** We report accuracy only. Confusion matrices, per-class recall, macro-F1, the false-alarm rate of the VLM alone, and baselines are still required. The alert-level benefit of FIRMS corroboration has not been measured, so we make no claim about false-positive reduction.
- **Tile-level output.** The model assesses a tile; it does not localize or measure burnt area.
- **Imagery latency.** Detection depends on Sentinel-2 revisit and cloud cover, and scars may be ploughed over before the next pass.

4.7. Responsible use

A false alert can harm a farmer through fines or loss of subsidy. The design therefore weights false alarms more heavily during training, exposes an `image_quality_limited` flag, labels uncorroborated alerts as pending verification, and logs every step so decisions can be reviewed. Given the accuracy reported above, RemoteSense should currently be used to prioritize field verification and never as sole evidence for enforcement.

5. Conclusion

The review of literature shows that active-fire products miss much crop residue burning, that scar-mapping methods return masks without field-level context, and that no fine-tuned vision language model had been applied to crop-residue burn scars. These gaps led to the selection of a LoRA-fine-tuned 450M-parameter VLM that reads Sentinel-2 RGB and SWIR composites. RemoteSense shows that such a model can be fine-tuned in about 16 minutes on a single GPU to produce structured burn-scar assessments, and that these assessments can be corroborated with FIRMS and routed as alerts through an end-to-end platform that also runs in the browser. Fine-tuning gives large gains over the base model on severity (11.8% to 53.9%) and mean five-field accuracy (21.9% to 74.8%), but binary burn detection (73.6%) is only modestly above the majority-class baseline, and baselines, field-verified labels and multi-season testing are still needed. Our next steps are to add NBR and CNN baselines, report full confusion matrices, expand and field-verify the labelled data, add multi-date inputs, and measure alert-level precision.

Availability. Code, model weights, ONNX export and evaluation notebooks are at <https://github.com/munish0838/parali>.

REFERENCES

- Anand, A., Imasu, R., Dhaka, S. K., & Patra, P. K. (2025). Domain adaptation and fine-tuning of a deep learning segmentation model of small agricultural burn area detection using high-resolution Sentinel-2 observations: A case study of Punjab, India. *Remote Sensing*, 17(6),

974.

- Ayanzadeh, A., Dixit, P., Kamal, S., & Halem, M. (2026). WildfireVLM: AI-powered analysis for early wildfire detection and risk assessment using satellite imagery (arXiv:2602.13305). arXiv.
- CEEW. (2025). How can satellite-borne remote sensing capture crop burning incidents? Council on Energy, Environment and Water, New Delhi, India. <https://www.ceew.in/publications/how-can-satellite-borne-remote-sensing-capture-crop-burning-and-farm-fires>
- Deshpande, M. V., Pillai, D., & Jain, M. (2022). Detecting and quantifying residue burning in smallholder systems: An integrated approach using Sentinel-2 data. *International Journal of Applied Earth Observation and Geoinformation*, 108, 102761.
- Detmers, T., Pagnoni, A., Holtzman, A., & Zettlemoyer, L. (2023). QLoRA: Efficient finetuning of quantized LLMs. In *Advances in Neural Information Processing Systems (NeurIPS)*.
- DPhi Tech. (n.d.). SimSat: Satellite imagery API. <https://dphi.tech>
- Drusch, M., et al. (2012). Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sensing of Environment*, 120, 25–36.
- European Space Agency. (n.d.-a). Φ -sat programme: Φ sat-2. ESA Φ -lab. <https://philab.esa.int/flagship-programmes/phi-sats-programme/>
- European Space Agency. (n.d.-b). Copernicus Data Space Ecosystem. <https://dataspace.copernicus.eu>
- Filippini, F. (2018). BAIS2: Burned area index for Sentinel-2. *Proceedings*, 2(7), 364.
- Hu, E. J., et al. (2022). LoRA: Low-rank adaptation of large language models. In *Proceedings of the International Conference on Learning Representations (ICLR)*.
- Key, C. H., & Benson, N. C. (2006). Landscape assessment: Ground measure of severity, the composite burn index. In *FIREMON: Fire Effects Monitoring and Inventory System* (USDA Forest Service General Technical Report RMRS-GTR-164-CD).
- Knopp, L., Wieland, M., Rättich, M., & Martinis, S. (2020). A deep learning approach for burned area segmentation with Sentinel-2 data. *Remote Sensing*, 12(15), 2422.
- Kuckreja, K., Danish, M. S., Naseer, M., Das, A., Khan, S., & Khan, F. S. (2024). GeoChat: Grounded large vision-language model for remote sensing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 27831–27840).
- Li, X., Wen, C., Hu, Y., Yuan, Z., & Zhu, X. X. (2024). Vision-language models in remote sensing: Current progress and future trends. *IEEE Geoscience and Remote Sensing Magazine*, 12(2), 32–66.
- Liquid AI. (2026). LFM2.5-VL-450M [Model card]. <https://huggingface.co/LiquidAI/LFM2.5-VL-450M>
- Miller, J. D., & Thode, A. E. (2007). Quantifying burn severity in a heterogeneous landscape with a relative version of the delta Normalized Burn Ratio (dNBR). *Remote Sensing of Environment*, 109(1), 66–80.
- NASA. (n.d.). Fire Information for Resource Management System (FIRMS). <https://firms.modaps.eosdis.nasa.gov>
- Ramo, R., et al. (2021). African burned area and fire carbon emissions are strongly impacted by small fires undetected by coarse resolution satellite data. *Proceedings of the National Academy of Sciences*, 118(9), e2011160118.
- Reuters. (2026, September 28). India's TakeMe2Space to launch orbital computing satellite on SpaceX rocket.
- Schroeder, W., Oliva, P., Giglio, L., & Csizar, I. A. (2014). The new VIIRS 375 m active fire detection data product: Algorithm description and initial assessment. *Remote Sensing of Environment*, 143, 85–96.
- Seidel, L., et al. (2025). Advancing early wildfire detection: Integration of vision language models with unmanned aerial vehicle remote sensing for enhanced situational awareness. *Drones*, 9(5).
- Soni, S., et al. (2025). EarthDial: Turning multi-sensory Earth observations to interactive dialogues. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 14303–14313).
- ThePrint. (2025). Are farm fires going uncouncted? ISRO study suggests farmers have figured out a way to evade satellites. <https://theprint.in/environment/are-farm-fires-going-uncouncted-isro-study-suggests-farmers-have-figured-out-a-way-to-evade-satellites/2797583/>
- The Tribune. (2025, December). Satellites miss majority of stubble fires, Delhi air pollution underestimated: Report [On the iFOREST Stubble Burning Status Report 2025]. <https://www.tribuneindia.com/news/delhi/satellites-miss-majority-of-stubble-fires-delhi-air-pollution-underestimated-report>
- Trigg, S., & Flasse, S. (2001). An evaluation of different bi-spectral spaces for discriminating burned shrub-savannah. *International Journal of Remote Sensing*, 22(13), 2641–2647.
- Walker, K., et al. (2022). Detecting crop burning in India using satellite data (arXiv:2209.10148). arXiv.
- Wang, F., et al. (2025). GeoLLaVA-8K: Scaling remote-sensing multimodal large language models to 8K resolution. In *Advances in Neural Information Processing Systems (NeurIPS)* (Vol. 38).