



iJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** X **Month of publication:** October 2026

DOI: <https://doi.org/10.22214/ijraset.2026.85433>

www.ijraset.com

Call: ☎ 08813907089

E-mail ID: ijraset@gmail.com

AI-Powered FRA Atlas: An AI- and WebGIS-Based Decision Support Framework for Forest Rights Act Implementation Monitoring

Mrs. Chandhana K S¹, Manish RZ², Prajwal H P³, Sagar G Kunte⁴, P B Darshan Prasad⁵

¹Assistant Professor, ^{2, 3, 4, 5}Seventh Semester Students, Department of Computer Science and Engineering, Jawaharlal Nehru New College of Engineering, Shivamogga, Visvesvaraya Technological University, Belagavi, Karnataka, India

Abstract: *The Forest Rights Act (FRA), 2006 recognizes the land and resource rights of forest-dwelling communities in India, but its implementation is hindered by paper-based records, fragmented databases, delayed claim verification and weak monitoring. This paper presents an AI-powered FRA Atlas integrated with a WebGIS-based Decision Support System (DSS) that combines digitized FRA records, Sentinel-2 satellite imagery, machine-learning land-cover classification, a retrieval-augmented policy assistant, a statutory scheme-eligibility rule engine and a tamper-evident audit trail in one platform. A Random Forest classifier operating on five Sentinel-2 bands and three spectral indices assigns pixels to seven land-cover classes. On a 4,200-sample holdout set it reaches 87.24% overall accuracy, Cohen's κ of 0.8511 and a macro F1-score of 0.8726, and 5-fold stratified cross-validation gives $87.35\% \pm 0.24\%$. The hybrid retrieval module achieves 85.71% top-1 accuracy and a mean reciprocal rank of 0.9286 on seven policy queries. The classifier is trained on calibrated synthetic spectral samples rather than field-labelled FRA land, so the reported figures measure separability under those distributions and are not field accuracy; validation on real labelled imagery remains the main open item.*

Index Terms: *Forest Rights Act, FRA Atlas, WebGIS, decision support system, remote sensing, land-cover classification, Random Forest, retrieval-augmented generation, Sentinel-2, e-governance.*

I. INTRODUCTION

The Forest Rights Act (FRA), 2006 was enacted to address historical injustice faced by forest-dwelling Scheduled Tribes and other traditional forest dwellers by recognizing their rights over forest land and resources. Implementation passes through several stages: claim submission by the Gram Sabha, verification, approval and documentation of titles for Individual Forest Rights (IFR), Community Rights (CR) and Community Forest Resource (CFR) rights.

In practice these stages rely on paper records and disconnected systems, which produce inconsistent data, delays and limited transparency. Discrepancies also arise between recorded claims and actual ground conditions. Meanwhile, GIS, remote sensing and AI now make it possible to classify land, detect change and visualize spatial data at scale.

The Ministry of Tribal Affairs (MoTA) has recognized this need. It placed a problem statement in Smart India Hackathon 2025 for an AI-powered FRA Atlas and WebGIS-based DSS to monitor IFR, CR and CFR rights [16], and the Northeast's first Digital FRA Atlas has been launched in Tripura [17]. This paper addresses that problem statement by describing and evaluating an integrated full-stack prototype that combines digitization, geospatial analysis, AI-based land classification and decision-support dashboards.

The primary contribution of this work is not a new classification algorithm. It is the integration of established methods and open technologies (OCR, retrieval-augmented generation, remote-sensing indices, Random Forest, FastAPI, a Next.js WebGIS front end) into a single system for FRA monitoring, together with a quantitative evaluation of its machine-learning and retrieval components and a clear statement of what that evaluation does and does not show.

II. MOTIVATION AND PROBLEM DEFINITION

Manual FRA processes are time-consuming and cannot provide real-time insight. Authorities need to track claim progress, detect inconsistencies and decide in time, yet data sits in departmental silos and paper files. Reported implementation gaps include high claim-rejection rates and low awareness among eligible households [10], [11]. Scanned claim forms and survey sheets are rarely searchable, and boundaries are often held as sketches, KML files or shapefiles that are not linked to claim status. The proposed system addresses the following objectives:

- 1) Design and develop an AI-powered FRA Atlas for monitoring land claims.
- 2) Integrate WebGIS for spatial visualization of FRA data.
- 3) Digitize scanned claim documents and extract structured fields.
- 4) Automate tracking of claim status and land distribution.
- 5) Improve verification using satellite imagery and AI-based analysis.
- 6) Provide dashboards, scheme-eligibility advice and a question-answering assistant for decision-makers.
- 7) Support transparency and accountability while respecting community data governance.

The resulting system is a computational research prototype that assists, but does not replace, the statutory authorities (Gram Sabha, Sub-Divisional and District Level Committees) that decide FRA claims.

III. RELATED WORK

A. FRA Implementation Studies

Singh studied FRA implementation in Sehore district, Madhya Pradesh, and found that ambiguities in the Act hindered the executing agencies [10]. Narayan analyzed the legal framework, noting implementation delays and the central role of the Gram Sabha [11]. Nikawat and Tak examined governance gaps through policy review and case studies, concluding that opacity, weak monitoring and data inconsistency are the main obstacles [2]. A Karnataka study based on field surveys reported significant delays in claim processing and inconsistencies in land-allocation records [3]. These studies are largely qualitative and propose no technical monitoring mechanism.

B. Existing Digital Initiatives

Manimegala et al. proposed an AI-powered FRA Atlas with a WebGIS DSS that applies machine learning to satellite imagery and combines it with government records [1]. State efforts such as the Maharashtra FRA Portal and the Tripura Digital FRA Atlas show that digitization is feasible [16], [17]. Critical commentary warns that AI-enabled FRA platforms need community consent and data-governance safeguards, because the Act itself has no data provisions [18].

C. Remote Sensing and Land-Cover Mapping

Rogan and Chen reviewed remote sensing for mapping land-cover and land-use change [4]. Drusch et al. described the Sentinel-2 mission, which provides free multispectral imagery at 10 m resolution [13]. Brown et al. presented Dynamic World, a near-real-time global 10 m land-cover dataset [14], and Zanaga et al. released the ESA WorldCover 10 m product [32]. Spectral indices such as NDVI [19], NDWI [20] and NDBI [21] are widely used to separate vegetation, water and built-up surfaces.

D. Machine Learning for Classification

Zhang et al. surveyed deep learning for remote sensing data and showed gains over traditional classifiers [5]. Ronneberger et al. introduced U-Net for pixel-wise segmentation [12], and Helber et al. released the EuroSAT Sentinel-2 benchmark [31]. Breiman's Random Forest offers robust ensemble classification [6] and is available in scikit-learn [30]. Devlin et al. introduced BERT for text understanding [9].

E. Document Digitization and Retrieval-Augmented Generation

Smith described the Tesseract OCR engine, an open-source system for recognizing printed text from scanned pages [22]. Lewis et al. introduced retrieval-augmented generation, in which a language model is supplied with retrieved passages so that answers are grounded in a knowledge base [23]. Together these allow scanned FRA documents and legal texts to be converted into searchable, queryable knowledge.

F. Platforms and Decision Support

Gorelick et al. presented Google Earth Engine for planetary-scale geospatial analysis [7]. Peng and Tsou described architectures for Internet GIS [15]. Shim et al. traced the evolution of decision-support technology and its role in organizational decision-making [8]. Although these works show the potential of each technology, few combine claim records, document digitization, imagery analysis, scheme-eligibility reasoning and decision dashboards for FRA specifically. The proposed framework focuses on this system-level integration. Table I summarizes the surveyed literature.

TABLE I
SUMMARY OF LITERATURE SURVEY

Ref	Focus	Key finding	Gap
[1]	AI FRA Atlas + WebGIS DSS	Unified, real-time monitoring	Data quality, integration
[2], [10], [11]	FRA policy / legal studies	Delays, opacity, inconsistent data	No technical solution
[3]	FRA in Karnataka	Claim delays, record mismatch	Regional, manual
[4], [13], [14]	Remote sensing	Scalable land-cover monitoring	Not linked to claims
[19]–[21]	Spectral indices	Simple vegetation, water, built-up cues	Need calibration per region
[5], [6], [12]	Deep learning, RF	High classification accuracy	Needs labels, compute
[22], [23]	OCR RAG	Searchable, grounded text	Quality of scans, grounding
[7]	Google Earth Engine	Cloud-scale processing	Needs connectivity
[8], [15]	DSS, Internet GIS	Data-driven decisions	Needs quality data
[9]	BERT	Strong text understanding	Not applied to FRA
[16]	Algorithmic governance	Consent and data risks	— only

IV. PROPOSED FRAMEWORK

The framework consists of nine components: (1) data ingestion and digitization of FRA records; (2) a geospatial database linking claims to boundaries; (3) Sentinel-2 acquisition and preprocessing; (4) AI-based land-cover classification; (5) inconsistency detection; (6) claim-status tracking and analytics; (7) a retrieval-augmented assistant and a statutory scheme-eligibility rule engine; (8) a tamper-evident audit trail; and (9) WebGIS visualization and decision dashboards. The overall processing flow is shown in Fig. 1, and results are presented to authorities through the FRA Atlas web application.

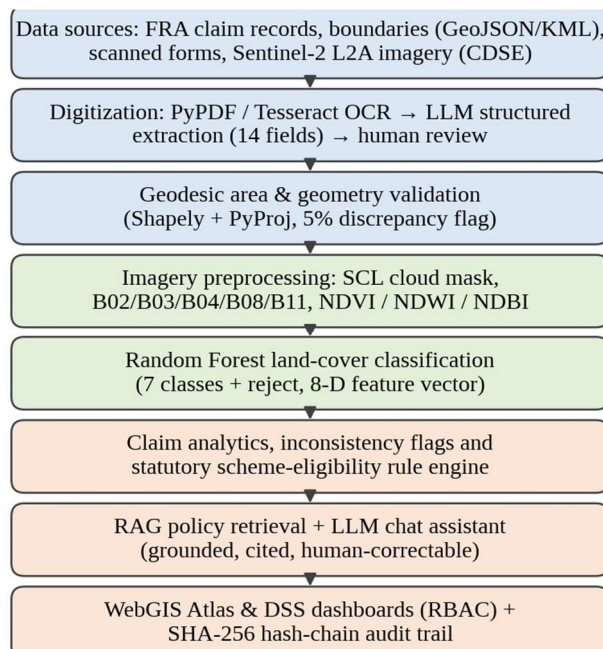


Fig. 1. Processing flow of the FRA Atlas and WebGIS DSS.

V. TECHNOLOGY STACK

The system is a full-stack web and geospatial application with a Python backend and a Next.js frontend. Table II lists the stack by layer and Fig. 2 shows how the layers relate.

TABLE II
TECHNOLOGY STACK OF THE FRA ATLAS

Layer	Technologies	Role
Frontend	Next.js 14, React 18, TypeScript, Tailwind CSS	Web app, routing, styling
Mapping / UI	Leaflet.js, Recharts, Lucide icons	Maps, charts, icons
Backend	Python 3.12, FastAPI, Uvicorn	Asynchronous REST API
Data	SQLAlchemy, Pydantic, SQLite (dev), PostgreSQL, PostGIS (prod)	ORM, validation, spatial storage
Security	JWT (HS256), bcrypt, role-based access control	Authentication, authorization
GIS	Shapely, PyProj (WGS84), GeoJSON / KML	Geometry, geodesic area
AI / OCR	Tesseract, PyPDF, Pillow, Groq LLM cascade, Gemini, regex	Digitization, extraction
ML	scikit-learn Random Forest	Land-cover classification
RAG / DSS	Hashed n-gram embeddings, cosine + keyword retrieval, rule engine	Policy grounding, eligibility
Remote sensing	Copernicus Sentinel-2 L2A via CDSE, Sentinel Hub API	Imagery, NDVI/NDWI/NDBI
Audit	SHA-256 hash chain	Tamper-evident logging

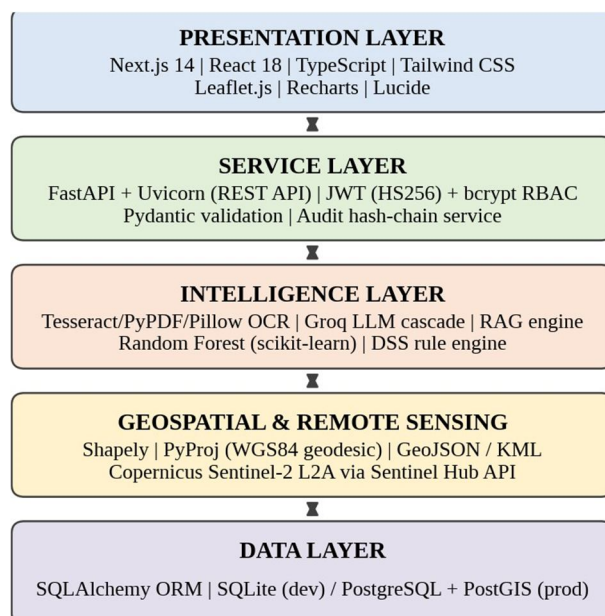


Fig. 2. Layered architecture and technology stack.

A. Frontend

The web client is built with Next.js 14 and React 18 in TypeScript, giving type-checked components and server-side rendering support. Tailwind CSS provides utility-first styling. Interactive maps are rendered with Leaflet.js, a lightweight open-source mapping library [26]. Recharts draws the dashboard charts and Lucide supplies icons.

B. Backend

The backend uses Python 3.12 with FastAPI served by Uvicorn. FastAPI offers asynchronous request handling, automatic OpenAPI documentation and tight integration with Pydantic for request and response validation [25]. SQLAlchemy provides the object-relational mapping between Python models and the database.

C. Data Layer

PostgreSQL with the PostGIS extension [27] is the production database, storing claim boundaries as geometry columns with spatial indexing, intersection and area queries. SQLite database files are used during local development for convenience.

D. Authentication and Security

Users authenticate with JSON Web Tokens (JWT) [28] signed with HS256; access tokens expire after 24 hours and refresh tokens after 7 days. Passwords are hashed with bcrypt and are never stored in plain text. Role-based access control defines six roles: Admin, State Officer, District Officer, Field Officer, Analyst and Citizen, restricting which states, districts and operations each user may access.

E. GIS Processing

Shapely performs geometric validation, including topology repair of self-intersections through a zero-width buffer, and PyProj handles coordinate transformations. Areas are computed geodesically on the WGS84 ellipsoid with pyproj.Geod rather than by planar approximation. The platform reads GeoJSON [29] (Polygon, MultiPolygon, Feature, FeatureCollection) and KML, and returns the centroid and bounding box of each boundary.

F. AI, OCR and Document Processing

PDF files are read with PyPDF. Scanned images are converted to grayscale, contrast-enhanced and sharpened with Pillow, then passed to Tesseract OCR [22]. A three-stage extractor converts raw text into 14 fields (claim ID, applicant name, father/husband name, age, gender, village, block, district, state, claim type, area, survey number, land use and application date): a Groq-hosted LLM [24] returning structured JSON is tried first, Google Gemini 1.5 Flash second, and a deterministic regular-expression parser is the final fallback. Low-confidence fields are routed to human review.

G. Remote Sensing

Sentinel-2A/B Harmonized L2A imagery is retrieved from the Copernicus Data Space Ecosystem through the Sentinel Hub Process API using OAuth2 client credentials [13]. The query covers a rolling 365-day window, and pixels are cloud-masked with the Scene Classification Layer (SCL), which removes seven classes: no data, saturated, cloud shadows, low/medium/high-probability cloud and cirrus. Bands B02, B03, B04 and B08 (10 m) and B11 (20 m) are used, and three indices are computed per boundary and date:

$$NDVI = (B08 - B04) / (B08 + B04) \quad (1)$$

$$NDWI = (B03 - B08) / (B03 + B08) \quad (2)$$

$$NDBI = (B11 - B08) / (B11 + B08) \quad (3)$$

where B03, B04, B08 and B11 are surface reflectance in the green, red, near-infrared and SWIR-1 bands [19]–[21]. Five raster layers are generated for each claim: true-colour RGB, colour infrared, and classified NDVI (4-class), NDWI (3-class) and NDBI (3-class) maps.

H. Deployment

The application can be packaged with Docker and Docker Compose, which start the frontend, backend and PostGIS database together so that development and deployment environments match.

VI. METHODOLOGY

A. Data Sources and Digitization

The system integrates (i) government FRA records (IFR, CR and CFR claims and their status), (ii) boundary and survey data in GeoJSON or KML form, (iii) Sentinel-2 imagery from the Copernicus Data Space Ecosystem, and (iv) scanned claim documents and field reports. The digitization pipeline has four steps: image cleanup, OCR, structured field extraction and human review of low-confidence fields.

B. Geospatial Database and Geo-referencing

Claim records are stored with spatial geometry, organized by state, district, block and village. Uploaded boundaries are validated, linked to claim records, and their geodesic area is compared with the claimed area. Table III shows the principal data entities.

TABLE III
PRINCIPAL DATA ENTITIES

Entity	Representative attributes
User	id, name, role (six RBAC roles), hashed password
Claim	id, claimant, village, rights type (IFR / CR / CFR), area, status, dates
Boundary	claim id, geometry (polygon), centroid, bounding box, source file
Document	claim id, file path, OCR text, extracted fields, per-field confidence
LandCover	boundary id, date, class fractions, NDVI / NDWI / NDBI, model version
Alert	claim id, type (area mismatch, land-use conflict), severity, review state
AuditBlock	previous hash, user, action, entity, timestamp, new value, hash

C. Land-Cover Classification

A Random Forest classifier [6] (scikit-learn [30]) with 60 trees, maximum depth 12, minimum 4 samples per split and a fixed random seed of 42 labels each pixel using an 8-dimensional feature vector: the five bands B02, B03, B04, B08, B11 and the indices of Eqs. (1)–(3). Seven classes are trained: forest/tree canopy, crop/agriculture, water, building/settlement, bare land/soil, grassland/pasture and road/compacted surface. An eighth label, *other*, is assigned to any pixel whose maximum class probability falls below a confidence threshold of 0.30, so uncertain pixels are rejected rather than forced into a class.

Training data. Labelled field samples for FRA landscapes were not available to the project. The classifier is therefore trained on 21,000 synthetic spectral samples (3,000 per class) drawn from surface-reflectance distributions calibrated against published references: ESA WorldCover 10 m [32], Dynamic World [14], EuroSAT [31] and the USGS/ASTER spectral library. The building class is modelled with two sub-clusters (concrete/tin and terracotta tile), and bare land follows an iron-oxide signature ($B04 > B03 > B02$). This design choice shapes how the results in Section VIII must be read.

D. Inconsistency Flagging

A claim is flagged when the declared area and the computed polygon area differ by more than 5%, or when the recorded land use conflicts with the observed land cover, for example a claim recorded as cultivation but classified mostly as forest. Flags are advisory and prompt field verification.

E. Claim Tracking and Analytics

The system tracks each claim through submission, verification, approval and title issue, and computes indicators such as pendency, rejection rate and area vested per district.

F. Retrieval-Augmented Assistant

Four official scheme guideline documents are split into 400-character passages with an 80-character (20%) overlap. Each passage is mapped to a 64-dimensional L2-normalized vector by deterministic character n-gram and trigram feature hashing. A query is scored against passages with a hybrid function, $score = 0.7 \cdot \cos(q, d) + 0.3 \cdot keyword-overlap(q, d)$ (4) and the top three passages are returned with source and page citations. The assistant is built on a Groq-hosted LLM cascade [24] (GPT-OSS 120B, Qwen, LLaMA 3.3 70B, LLaMA 3.1 70B) with temperature 0.2, a 1,800-token limit and a context of the last six turns. Its system prompt is injected with the claim record, satellite land-cover statistics and rule-engine output so that answers are grounded in system data; if no LLM is reachable a deterministic rule-based responder is used.

G. Scheme-Eligibility Rule Engine

A deterministic engine evaluates six schemes: PM-KISAN, PMKSY, PM Van Dhan Vikas Yojana, PMAY-G, MGNREGA FRA land development and Jal Jeevan Mission. Inputs are the FRA claim (title status, rights type, area), satellite land-cover fractions (crop, forest, water, building, bare land) and detected assets.

Each scheme returns one of three states, *eligible*, *conditional* or *ineligible*, with a score. For PM-KISAN, for example, a parcel at least 75% under water is ineligible, a pending title is conditional, and an approved title with crop cover above 15% (or a detected farm asset) is eligible. Because the rules are explicit, every recommendation can be traced to a statutory condition.

H. Tamper-Evident Audit Trail

Every state-changing action (user creation, claim changes, status changes, document uploads, scheme evaluations) is appended to a hash chain:

$$H_n = \text{SHA-256}(H_{n-1} \parallel \text{user} \parallel \text{action} \parallel \text{entity} \parallel \text{id} \parallel \text{time} \parallel \text{value}) \quad (5)$$

with a genesis hash of 64 zeros [34]. Re-computing the chain detects any altered or deleted record. This is blockchain-inspired but is a single-operator log, not a distributed ledger.

I. Decision Support and Visualization

The DSS layer follows established decision-support principles [8]. The WebGIS front end [15] provides layered maps, filters by state, district and rights type, and dashboards with charts and tables. Role-based views serve different levels of authority.

VII. SYSTEM DESIGN AND IMPLEMENTATION

The FastAPI application is organized into routers for authentication, claims, boundaries, documents, remote-sensing analysis, analytics, DSS evaluation and the assistant, with Pydantic schemas validating every request and response. The Next.js application provides a login page, a map view with claim layers and land-cover overlays, a claims register with filters, a document-upload page showing OCR results for review, an analytics dashboard and the assistant chat panel.

In a typical workflow an officer logs in and uploads a scanned claim form and its boundary file. The backend runs OCR and extraction, validates the geometry and computes its geodesic area, and requests a Sentinel-2 analysis for the polygon. The officer reviews the extracted fields, and the map shows the claim with its land-cover layers and any flags. For a chat query the backend resolves the claim, fetches land-cover statistics, retrieves policy passages, evaluates scheme eligibility and passes this evidence to the LLM, as the response pipeline in Fig. 3 shows.

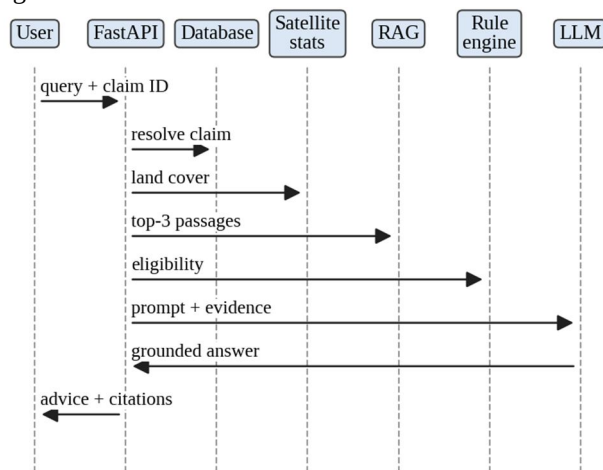


Fig. 3. Chat response pipeline: evidence is gathered from the database, satellite statistics, RAG and rule engine before the LLM is called.

VIII. EXPERIMENTAL RESULTS

A. Experimental Setup

All figures in this section were produced by evaluation scripts run against the implemented classifier, retrieval module and rule engine. The classifier was assessed with 5-fold stratified cross-validation (shuffled, seed 42) over all 21,000 samples, and separately with an 80/20 stratified holdout (16,800 training and 4,200 test samples, 600 per class). Metrics are overall accuracy, per-class precision (P), recall (R) and F1, macro- and weighted-averaged F1, and Cohen's κ [33]:

$$F1 = 2PR / (P + R) \quad (6)$$

$$\kappa = (p_o - p_c) / (1 - p_c) \quad (7)$$

where p_o is observed and p_c chance agreement. Retrieval was assessed by top-1 accuracy and mean reciprocal rank, $MRR = (1/|Q|) \sum 1/\text{rank}_i$.

B. Cross-Validation

Table IV and Fig. 4 give the fold-wise results. Accuracy ranges from 0.8705 to 0.8764 with a mean of 0.8735 and a standard deviation of 0.0024, and macro F1 follows it closely (mean 0.8738, $\sigma = 0.0023$). Mean macro precision and recall are 0.8744 and 0.8735. The narrow spread shows that the result does not depend on a particular split of the synthetic sample.

TABLE IV
5-FOLD STRATIFIED CROSS-VALIDATION (n = 21,000)

Metric	Mean	Std.	F1	F2	F3	F4	F5
Accuracy	0.8735	0.0024	0.8731	0.8705	0.8762	0.8764	0.8714
F1 (macro)	0.8738	0.0023	0.8732	0.8710	0.8763	0.8767	0.8717

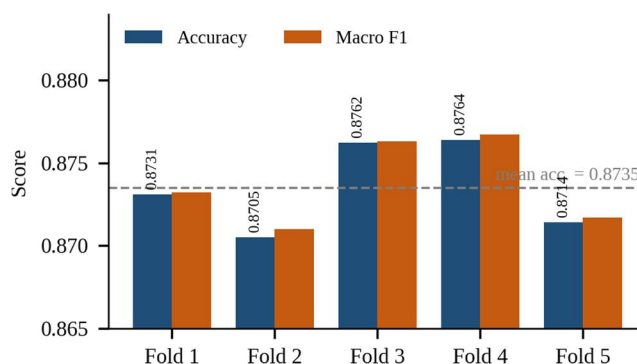


Fig. 4. Fold-wise accuracy and macro F1 from 5-fold stratified cross-validation.

C. Holdout Evaluation and Per-Class Performance

On the 4,200-sample holdout set the classifier attains an overall accuracy of 0.8724, Cohen's κ of 0.8511, and macro and weighted F1 of 0.8726. The holdout accuracy lies within 0.0011 of the cross-validation mean, so there is no sign of overfitting to a particular split. Per-class results are in Table V and Fig. 5

TABLE V
PER-CLASS CLASSIFICATION REPORT (HOLDOUT, n = 4,200)

Class	Precision	Recall	F1-score	Support
Forest	0.9288	0.9133	0.9210	600
Crop	0.8061	0.7900	0.7980	600
Water	1.0000	1.0000	1.0000	600
Building	0.9399	0.9117	0.9255	600
Bare Land	0.8264	0.8333	0.8299	600
Grassland	0.7803	0.7933	0.7868	600
Road	0.8304	0.8650	0.8473	600
Macro avg	0.8731	0.8724	0.8726	4,200

Fig. 5. Per-class precision, recall and F1 on the holdout set.

Water is classified perfectly ($F1 = 1.000$) because its strong NIR and SWIR absorption makes it spectrally unique. Building (0.9255) and forest (0.9210) exceed 0.92. The weakest classes are grassland (0.7868) and crop (0.7980), whose herbaceous reflectance overlaps.

D. Confusion Analysis

Fig. 6 shows the confusion matrix and Fig. 7 breaks the 536 holdout errors down by class pair. Three pairs account for 345 errors (64.4%): crop↔grassland (146, 27.2%), road↔bare land (105, 19.6%) and forest↔crop (94, 17.5%). Crop and grassland are both photosynthetic herbaceous covers with overlapping NDVI ranges; road and bare land share similar visible albedo and SWIR reflectance; and forest is separated from crop mainly by canopy height, which 2-D multispectral features do not capture. Water has no confusion with any class.

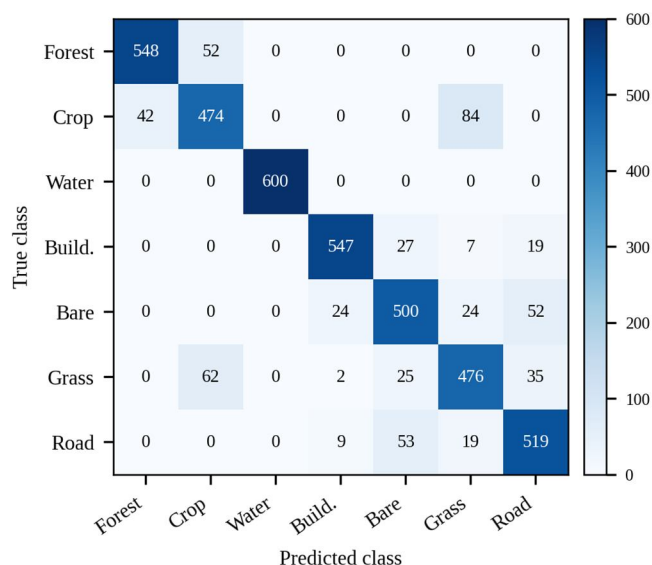


Fig. 6. Confusion matrix on the 4,200-sample holdout set (rows: true, columns: predicted).

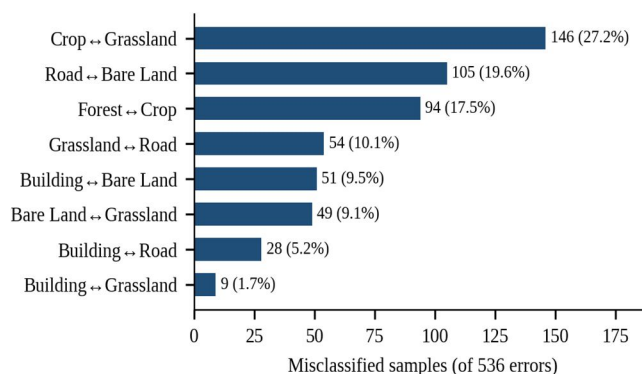


Fig. 7. Contribution of class pairs to the 536 holdout misclassifications (both directions summed).

E. Feature Importance

Fig. 8 reports mean-decrease-impurity importances. B04 (Red) ranks first at 0.1662, followed by NDVI (0.1598), B11 SWIR-1 (0.1579) and B08 NIR (0.1461).

The three engineered indices together contribute 38.5% of total importance (NDVI 16.0%, NDWI 11.3%, NDBI 11.2%), supporting their inclusion alongside raw bands. The green band contributes least (0.0574) since its range varies little across classes. Impurity-based importances are known to favour correlated and high-variance features, so they indicate usage within the model rather than causal relevance.

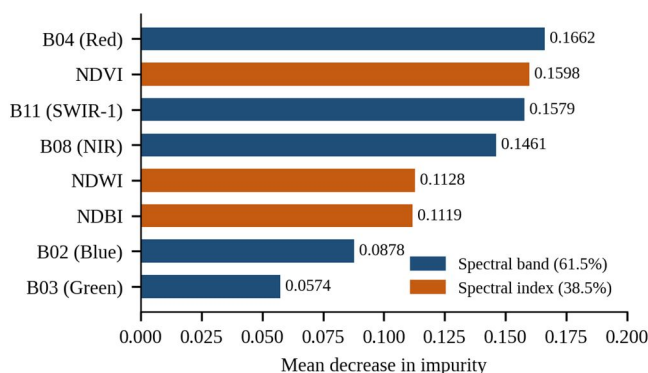


Fig. 8. Random Forest feature importance (mean decrease in impurity).

F. Retrieval Performance

The hybrid retriever was tested on seven queries over the four-document policy corpus, each with a known target scheme (Table VI). It returned the correct scheme at rank one for six queries, giving a top-1 accuracy of 85.71%. The single failure (Q7, a query naming PMAY-G in full) ranked the correct document second, so the MRR is $(6 \times 1 + 0.5)/7 = 0.9286$. Query 7 failed because “Pradhan Mantri” appears in several scheme names and the PMKSY document scored higher through shared terms; its top score of 0.1316 was also the lowest of all queries (Fig. 9), so a minimum-score threshold could route such queries to a clarification step. Hashed n-gram vectors capture lexical rather than semantic similarity, and a sentence-embedding model [35] is the natural next step. With only seven queries, these figures are indicative; one more miss would lower top-1 accuracy to 71.4%.

TABLE VI
RAG RETRIEVAL EVALUATION (7 QUERIES)

#	Query focus	Target	Retrieved	Score	OK
1	PM-KISAN for FRA gatta holders	PM-KISAN	PM-KISAN	0.4134	Yes
2	Micro-irrigation, water deficit	PMKSY	PMKSY	0.3340	Yes
3	Van Dhan, NTFP, tribal SHG	VDVY	VDVY	0.5102	Yes
4	PMAY housing grant, kutch/pucca	PMAY-G	PMAY-G	0.3407	Yes
5	Water availability, irrigation	PMKSY	PMKSY	0.4743	Yes
6	Forest canopy, minor forest produce	VDVY	VDVY	0.5375	Yes
7	Rural housing, Pradhan Mantri Awas	PMAY-G	PMKSY	0.1316	No

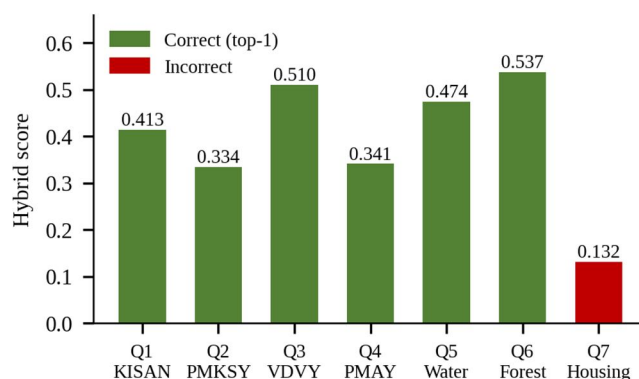


Fig. 9. Hybrid retrieval score of the top-ranked document for each query.

G. Document Extraction Confidence

Table VII lists the confidence assigned to each field by the rule-based fallback parser when its pattern matches, which ranges from 0.85 to 0.96; PDF text extraction is assigned 0.96, and LLM-based extraction reports 0.90–0.96. These are confidence scores built into the extractor and are used to route low-confidence fields to human review. They are **not** measured accuracies; character error rate and field-level accuracy against annotated FRA forms have not yet been measured.

TABLE VII
ASSIGNED CONFIDENCE OF RULE-BASED FIELD EXTRACTION

Field	Conf.	Field	Conf.
Claim type	0.96	Application date	0.92
Claim ID	0.95	Village	0.91
State	0.95	Father name	0.90
Age	0.95	Gender	0.90
District	0.94	Survey number	0.90
Area	0.93	Block	0.88
Applicant name	0.92	Land use	0.85

H. Summary of System Components

Table VIII consolidates the measured results and the implemented component properties.

TABLE VIII
SUMMARY OF RESULTS

Component	Metric	Value
Random Forest	CV accuracy (5-fold)	87.35% ± 0.24%
	Holdout accuracy	87.24%
	Cohen's κ	0.8511
	Macro F1	0.8726
	Best / weakest class	Water 1.000 / Grassland 0.787
RAG retrieval	Top-1 accuracy	85.71% (6/7)
	MRR	0.9286
DSS rule engine	Schemes / states	6 / 3
Remote sensing	Resolution; inputs	10 m; 5 bands + 3 indices
	Cloud mask	SCL, 7 classes
Audit chain	Hash; verification	SHA-256; full chain check

IX. DISCUSSION

The Random Forest baseline separates the seven classes well under the calibrated distributions, and its errors fall where spectral theory predicts: between herbaceous covers and between soil-like surfaces. These are also the distinctions that matter for FRA, since cultivation versus forest or grassland is the evidence used to test claim consistency, and the confidence-based reject class gives a way to defer such pixels to field verification.

The headline accuracy has to be read with care. Training and test samples come from the same synthetic generator, so the 87% figure measures separability of the modelled distributions. Real imagery adds atmospheric residuals, mixed pixels, terrain shadow, seasonal phenology and sensor noise, and real-world accuracy is expected to be lower. The cross-validation spread of 0.24% reflects the sampling variability of the synthetic data, not variability between landscapes. Validation on field-labelled Sentinel-2 pixels from the four target states is therefore the most important next step.

Beyond the classifier, the platform adds explainable components that are a better fit for statutory settings than a black-box score. The rule engine ties each scheme recommendation to explicit conditions, the RAG module returns cited policy passages, and the hash-chained audit log makes record tampering detectable. The choice of open-source components (FastAPI, PostGIS, Leaflet, Tesseract) keeps licensing cost low, which matters for adoption by state departments. Table IX contrasts the system with current practice.

TABLE IX
MANUAL PRACTICE VERSUS PROPOSED SYSTEM

Aspect	Manual / legacy practice	Proposed FRA Atlas
Records	Paper files, separate databases	Digitized, geo-referenced, single database
Digitization	Manual data entry	OCR + LLM extraction of 14 fields, human review
Land assessment	Field visits only	Sentinel-2 classification (87.24% on synthetic holdout) supporting field visits
Spatial check	Manual sketch comparison	Geodesic area with 5% discrepancy flag
Flexibility	Manual checklist	Rule engine with traceable conditions, LLM advisory
Policy lookup	Printed guidelines	RAG retrieval (85.71% top-1, n = 7)
Audit	Paper registers	SHA-256 hash chain
Access	Departmental	Role-based web access

The claim that processing time falls from weeks to minutes is a design expectation of the automated pipeline; end-to-end latency and officer task-completion time have not yet been benchmarked and no timing figure is claimed here.

X. PRIVACY, ETHICS AND GOVERNANCE

FRA data concerns vulnerable communities. The system therefore applies role-based access control, hashed credentials, token-based sessions, audit logging and a community-consent protocol, since the Gram Sabha holds statutory authority over claims [18]. Encrypted transport must be enforced at deployment. AI outputs are decision support only. Final decisions remain with the statutory committees, and flagged inconsistencies must not be used as automatic grounds for rejecting claims. LLM output must always be shown with its source passages and be open to human correction. Because the classifier is currently trained on synthetic data, its land-cover output must not be used as evidence in any real claim decision until it has been validated on field data.

XI. LIMITATIONS

- (1) The classifier is trained and tested on synthetic spectral samples; no accuracy on field-labelled imagery is reported.
- (2) Only a Random Forest was implemented and evaluated; a CNN/U-Net was considered but has not been built, and no comparison between them exists.
- (3) Classification is per pixel on 10 m and 20 m data, so small plots and mixed pixels are limited, and canopy height is unavailable.
- (4) The retrieval evaluation uses seven queries and four documents, and its embedding is lexical.
- (5) OCR character error rate, extraction field accuracy, inconsistency-flag precision and recall, API latency under load and officer usability have not been measured.
- (6) Quality of OCR falls on poor scans and handwriting, and LLM extraction may err and needs human review.
- (7) Satellite retrieval requires connectivity and CDSE credentials; the LLM cascade depends on external APIs.
- (8) The audit chain protects integrity within one deployment but is not a distributed ledger.
- (9) The current scope covers four states.

XII. FUTURE WORK

- 1) Validate and retrain the classifier on field-labelled Sentinel-2 pixels from Madhya Pradesh, Odisha, Telangana and Tripura, using spatially separated test regions.
- 2) Implement and benchmark a CNN/U-Net against the Random Forest; add temporal composites and change detection.
- 3) Replace hashed n-gram retrieval with sentence embeddings and evaluate on a larger expert-reviewed question set.
- 4) Measure OCR error rate, field-level extraction accuracy and flag precision/recall; run a usability study with officials.
- 5) Multilingual interface and OCR for regional languages; mobile field-verification app with offline support.
- 6) Higher-resolution imagery and drone data for small holdings; predictive analytics for claim pendency.
- 7) Integration with national FRA and tribal-welfare platforms; privacy-preserving data sharing across states.

XIII. CONCLUSION

This paper presented an AI-powered FRA Atlas with a WebGIS-based Decision Support System, covering its design, implementation and first quantitative evaluation. The platform combines a Next.js frontend, a FastAPI backend, geodesic GIS validation, OCR and LLM-based document extraction, Sentinel-2 land analysis, RAG-based policy retrieval, a statutory scheme-eligibility rule engine and a SHA-256 audit chain. The Random Forest land-cover classifier reached 87.24% holdout accuracy ($\kappa = 0.8511$, macro F1 = 0.8726) and $87.35\% \pm 0.24\%$ in 5-fold cross-validation, with errors concentrated between crop and grassland and between road and bare land, and the retriever reached 85.71% top-1 accuracy (MRR 0.9286). Because these results rest on calibrated synthetic training data and a small retrieval test set, the system should be regarded as a research and engineering prototype; validation on field-labelled imagery, measured OCR and usability results, and community-centred governance are required before real-world deployment.

XIV. ACKNOWLEDGMENT

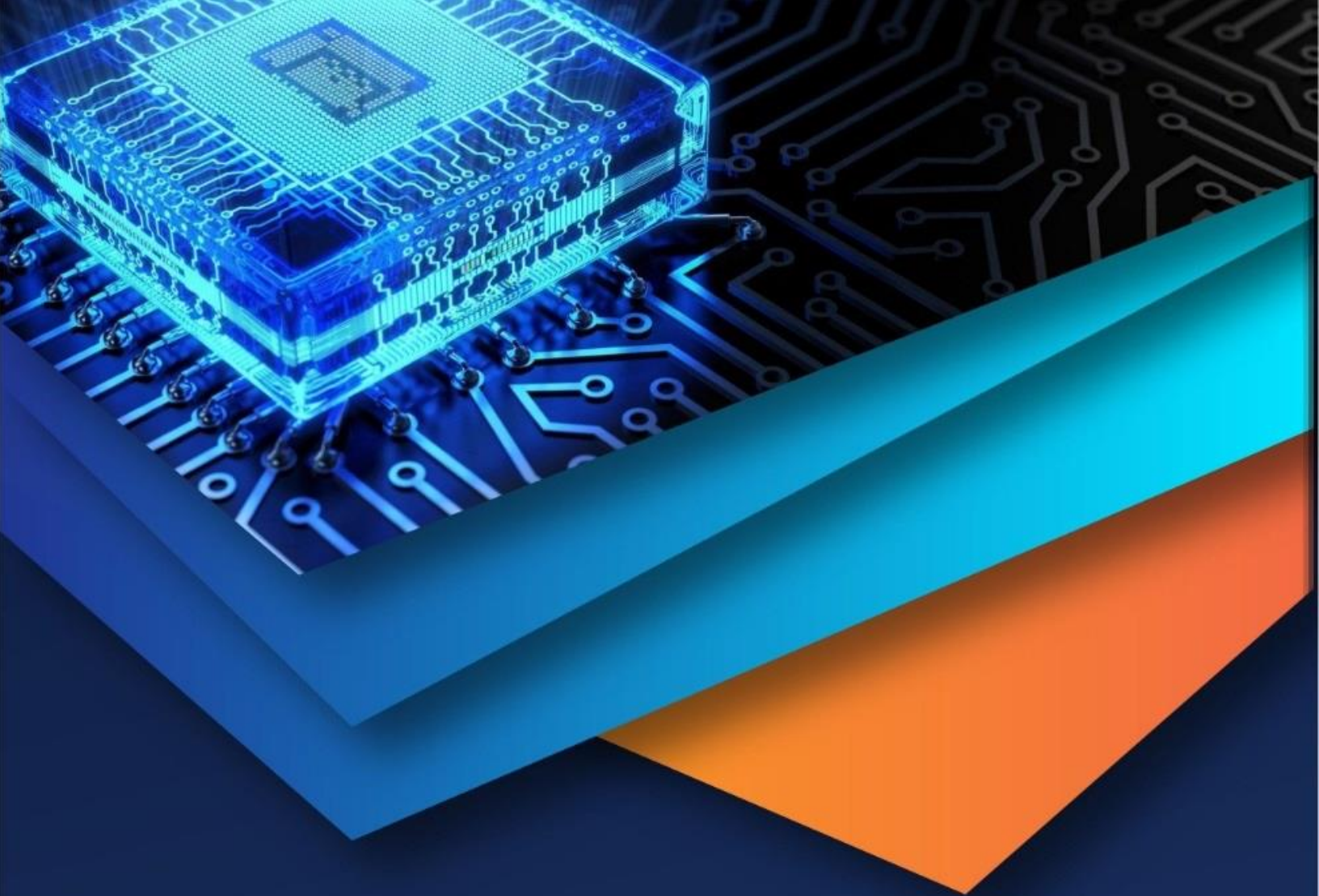
The authors thank Dr. Jalesh Kumar, Professor and Head, Department of Computer Science and Engineering, and Dr. Y. Vijaya Kumar, Principal, JNN College of Engineering, Shivamogga, for their support.

REFERENCES

- [1] M. Manimegala, T. Gayathri, M. M. Deva Harsar, and Y. Hans Rohit, "AI-Powered FRA Atlas and WebGIS-Based Decision Support System for Forest Rights Act Implementation Monitoring," 2026.
- [2] K. Nikawat and P. Tak, "A Critical Study of Forest Rights Act, 2006," 2024.
- [3] S. B. S., K. G. Dileep Kumar, and B. Ramesh, "An Analytical Study of Forest Rights Act Implementation in Karnataka," 2020.
- [4] J. Rogan and D. Chen, "Remote sensing technology for mapping and monitoring land-cover and land-use change," *Progress in Planning*, vol. 61, no. 4, pp. 301–325, 2004.
- [5] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 22–40, 2016.
- [6] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [7] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, "Google Earth Engine: Planetary-scale geospatial analysis for everyone," *Remote Sens. Environ.*, vol. 202, pp. 18–27, 2017.
- [8] J. P. Shim, M. Warkentin, J. F. Courtney, D. J. Power, R. Sharda, and C. Carlsson, "Past, present, and future of decision support technology," *Decision Support Systems*, vol. 33, no. 2, pp. 111–126, 2002.
- [9] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. NAACL-HLT*, 2019, pp. 4171–4186.
- [10] S. Singh, "The implications of the Scheduled Tribes and Other Traditional Forest Dwellers (Recognition of Forest Rights) Act, 2006," *Indian Forester*, vol. 140, no. 11, pp. 1081–1084, 2014, doi: 10.36808/if/2014/v140i11/54443.
- [11] S. Narayan, "Legal framework for forest rights in India: A critical analysis," Univ. of North Bengal, 2024. [Online]. Available: <https://ir.nbu.ac.in/handle/123456789/5580>
- [12] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*, LNCS vol. 9351, 2015, pp. 234–241.
- [13] M. Drusch *et al.*, "Sentinel-2: ESA's optical high-resolution mission for GMES operational services," *Remote Sens. Environ.*, vol. 120, pp. 25–36, 2012.
- [14] C. F. Brown *et al.*, "Dynamic World, near real-time global 10 m land use land cover mapping," *Scientific Data*, vol. 9, art. 251, 2022.
- [15] Z.-R. Peng and M.-H. Tsou, *Internet GIS: Distributed Geographic Information Services for the Internet and Wireless Networks*. Hoboken, NJ, USA: Wiley, 2003.
- [16] Ministry of Tribal Affairs, Govt. of India, "Forest Rights Act (FRA) digital platform," Lok Sabha unstarred question AU2117. [Online]. Available: https://eparlib.sansad.in/bitstream/123456789/3022375/1/AU_2117_r9gKnx.pdf
- [17] "Tripura Governor launches Northeast's first digital FRA atlas for tribal rights," NewKerala.com, Jul. 1, 2026. [Online]. Available: <https://www.newkerala.com/news/a/committed-tribal-dignitytripura-governor-launch-es-northeast-first-426.htm>
- [18] "Two faces of the same algorithm: Algorithmic disparate impact on India's Adivasi communities," LiveLaw. [Online]. Available: <https://www.livelaw.in/lawschool/articles/algorithmic-disparateimpact-india-adi-vasi-communities-544518>
- [19] J. W. Rouse, R. H. Haas, J. A. Schell, and D. W. Deering, "Monitoring vegetation systems in the Great Plains with ERTS," in *Proc. 3rd ERTS Symp.*, NASA SP-351, 1974, pp. 309–317.
- [20] S. K. McFeeters, "The use of the normalized difference water index (NDWI) in the delineation of open water features," *Int. J. Remote Sens.*, vol. 17, no. 7, pp. 1425–1432, 1996.
- [21] Y. Zha, J. Gao, and S. Ni, "Use of normalized difference built-up index in automatically mapping urban areas from TM imagery," *Int. J. Remote Sens.*, vol. 24, no. 3, pp. 583–594, 2003.
- [22] R. Smith, "An overview of the Tesseract OCR engine," in *Proc. 9th Int. Conf. Document Analysis and Recognition (ICDAR)*, 2007, pp. 629–633.
- [23] P. Lewis *et al.*, "Retrieval-augmented generation for knowledge-intensive NLP tasks," in *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp. 9459–9474.
- [24] Groq Inc., "GroqCloud API documentation." [Online]. Available: <https://console.groq.com/docs>
- [25] S. Ramírez, "FastAPI documentation." [Online]. Available: <https://fastapi.tiangolo.com>
- [26] V. Agafonkin, "Leaflet: An open-source JavaScript library for interactive maps." [Online]. Available: <https://leafletjs.com>
- [27] PostGIS Project, "PostGIS documentation." [Online]. Available: <https://postgis.net>
- [28] M. Jones, J. Bradley, and N. Sakimura, "JSON Web Token (JWT)," IETF RFC 7519, May 2015.



- [29] H. Butler, M. Daly, A. Doyle, S. Gillies, S. Hagen, and T. Schaub, "The GeoJSON format," IETF RFC 7946, Aug. 2016.
- [30] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.
- [31] P. Helber, B. Bischke, A. Dengel, and D. Borth, "EuroSAT: A novel dataset and deep learning benchmark for land use and land cover classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 7, pp. 2217–2226, 2019.
- [32] D. Zanaga *et al.*, "ESA WorldCover 10 m 2020 v100," Zenodo, 2021, doi: 10.5281/zenodo.5571936.
- [33] J. Cohen, "A coefficient of agreement for nominal scales," *Educ. Psychol. Meas.*, vol. 20, no. 1, pp. 37–46, 1960.
- [34] NIST, "Secure Hash Standard (SHS)," FIPS PUB 180-4, Aug. 2015.
- [35] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence embeddings using Siamese BERT-networks," in *Proc. EMNLP-IJCNLP*, 2019, pp. 3982–3992.



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)