An Adaptive and Modern Land Use / Land Cover Classification System for Indonesia Using Multisensor Earth Observation Imagery and Data-Driven Techniques: Collecting Training Data with Dashcam Camera

Hadi F. [1*], Wahyuddin Y.[1], Sabri L.M. [1,] , Suprayogi A.[1], and Ramdani F [2].

[1]Lecturer, Department of Geodetic Engineering, Faculty of Engineering, Universitas Diponegoro, Indonesia

[2] Assistant Professor, Graduate School of Humanities and Social Science, University of Tsukuba, Japan

*firmanhadi@ft.undip.ac.id

Abstract *The lack of publicly available datasets for land use and land cover (LULC) training samples in Indonesia poses considerable difficulties in verifying historical image classification outcomes. These necessitate expensive field surveys that are often beyond financial reach.. This study addresses the critical need for accessible and cost-effective LULC training data collection by developing an innovative system that utilizes consumer-grade dashcams with Global Positioning System (GPS) capabilities to generate georeferenced video datasets with embedded coordinate and timestamp information. A comprehensive Python application was developed with assistance from Claude.ai within the Google Colab environment to automate the conversion of dashcam videos to individual images, perform optical character recognition for text detection, store the extracted data in a structured database, and deploy the system publicly using pyngrok for collaborative access. The application comprises four essential functionalities: camera information correction, land cover labeling, coordinate adjustment for sample locations, and data export capabilities in Comma-Separated Values (CSV) and GeoJSON formats. Analysis of 20,819 test images revealed that 6,425 photographs (30.86 percent) required manual verification and correction of camera-derived information, underscoring the need for quality control mechanisms in automated data collection systems. The web-accessible interface enables collaborative data processing by multiple users, significantly accelerating the correction and labeling workflow compared with traditional field-based approaches. The integrated dashboard provides end users with spatial filtering and area selection tools for customized dataset downloads, thereby enhancing data accessibility across diverse research applications. This prototype system serves as a foundational component for establishing a comprehensive, scalable, and cost-effective land cover classification framework for Indonesia, offering substantial improvements in data-collection efficiency and collaborative research capabilities for Earth observation applications.*

Keywords: land cover classification, dashcam imagery, collaborative mapping, Indonesia datasets, automated processing

1. Introduction

Land use and land cover (LULC) information is one of the most vital components in environmental monitoring, natural resource management, and sustainable development planning. In Indonesia, a country with over 17,000 islands and diverse ecosystems, including tropical rainforests, agricultural areas, urban settlements, and coastal zones, accurate and current LULC data is increasingly essential for informed decision-making at local, regional, and national levels (Kelly-Fair et al., 2022; Saputra & Lee, 2019). The constantly changing landscape of Indonesia, affected by rapid urbanization, agricultural development, deforestation, and climate change, requires ongoing monitoring and classification systems that can fully capture these changes over space and time (Ambarwulan et al., 2022; Sari et al., 2023). Recent studies show that LULC products serve as the fundamental thematic information needed for monitoring, resource management, planning, and accurate simulation of future scenarios, which directly supports Indonesia's Sustainable Development Goals (SDGs), especially SDG 11 (Sustainable Cities and Communities) and SDG 15 (Life on Land) (Kelly-Fair et al., 2022; Saputra & Lee, 2019).

Remote sensing technologies have become some of the most popular data sources for LULC mapping because they can extract information from large geographic areas at different spatial and temporal resolutions. Earth observation satellites, including Landsat, Sentinel, and various commercial platforms, produce large amounts of imagery that support systematic landscape monitoring (Fikriyah, 2020; Sari et al., 2023). However, the success of image classification algorithms, whether traditional pixel-based methods or modern machine learning techniques, depends on having reliable training samples and validation datasets (Sukojo & Ramadaningtyas, 2025). These training samples provide the ground truth data needed to help classification algorithms distinguish between different land cover types and land use categories. Without enough and representative training data, even the most advanced classification algorithms cannot generate accurate and dependable LULC maps (Utomo et al., 2025).

Traditional ground truth data collection methods require teams of surveyors to visit sample locations, record observations, capture photographs, and document coordinates using specialized GPS equipment. This conventional approach, while producing high-quality data, involves significant financial investments in personnel, transportation, equipment, and time (Miranda & Aryuni, 2021; Suni et al., 2023). Landscape complexity further increases these challenges, as classification accuracy heavily depends on data selection and validation strategies, with studies reporting overall accuracies ranging from 70% to 90% depending on the approaches (2022; Sukojo & Ramadaningtyas, 2025; Utomo et al., 2025). The comparison between manual on-screen digitizing (87.2% overall accuracy) and deep learning methods (70.2% accuracy) highlights ongoing challenges in training due to label ambiguity and class definitions for automated systems (Sukojo & Ramadaningtyas, 2025). For many researchers, academic institutions, and government agencies operating under budget restrictions, conducting extensive field campaigns across multiple provinces or even within a large region becomes prohibitively expensive and logistically difficult (Miranda & Aryuni, 2021; Suni et al., 2023).

Furthermore, the time factor complicates LULC validation. When verifying past satellite classifications or conducting multi-temporal change studies, ground conditions during current visits might not match historical landscapes (Yuliani & Ramli, 2024). Historical LULC samples are seldom stored in public repositories, making it hard to validate older imagery (Luvai et al., 2021). This restricts the ability to assess classification accuracy, perform temporal analysis, or develop machine learning models that need extensive multi-temporal training data (Heryadi & Miranda, 2020; Luvai et al., 2021).

Indonesia lacks publicly available LULC training datasets, a key obstacle for Earth observation research and applications. Unlike large benchmarks like ImageNet in computer vision, the remote sensing field lacks similar resources, hindering model transferability and comparison (Li et al., 2020). While crowdsourced data can form large-scale benchmarks—one study created a 24,000+ image dataset annotated with OpenStreetMap—Indonesia's infrastructure for systematic data collection, quality control, archiving, and distribution is absent. This affects early-career researchers, students, and under-resourced practitioners who rely on such data for classification, thesis work, or mapping (Miranda & Aryuni, 2021; Suni et al., 2023).

Recent advances in consumer electronics and mobile computing enable new methods of data collection (Liu et al., 2015; Zhao et al., 2022). Modern dashboard cameras equipped with GPS, initially designed for vehicle security, now incorporate geographic data into videos (Joshi et al., 2025). Studies indicate that consumer devices are effective for geospatial data collection, with mobile crowdsensing frameworks inferring unsensed data and integrating vehicle sensing (Zhao et al., 2022). Smartphones and action cameras also support georeferenced image collection during fieldwork (Liu et al., 2019; Wang et al., 2015).

The integration of AI tools and automated workflows improves consumer data collection. Research indicates that consumer dashcam videos can generate accurate geospatial measurements, with pipelines achieving average geolocation errors of about 2.83 meters and height errors of 2.09 meters for trees and 0.88 meters for poles (Joshi et al., 2025). OCR extracts embedded coordinates from video frames, and Python develops pipelines to convert raw footage into structured data (Joshi et al., 2025). Crowdsourced geodata from GPS devices can annotate remote sensing imagery at scale, expanding datasets through public mobile and GIS feeds (Li et al., 2020). Web platforms enable distributed teams to perform data quality control and validation without requiring specialized software or training (Liu et al., 2019; Zhao et al., 2022).

These technological convergences create opportunities to democratize geospatial data collection and engage broader communities in Earth observation activities. Research in sparse mobile crowdsensing (SMCS) has shown that methods can select small subsets of locations for sensing and then infer unsensed areas, reducing costs while utilizing spatial and temporal correlations in urban sensing data (Zhao et al., 2022). Unmanned aerial vehicles (UAVs) demonstrate the potential for consumer-grade technology to produce very high-resolution ortho-mosaics with ground sampling distances of about 2.89 cm/pixel and generate reliable LULC maps with acceptable accuracy coefficients (Wang et al., 2015). The integration of crowdsourced annotation methods using OpenStreetMap points and vectors has proven successful in creating globally distributed benchmarks for remote sensing image classification (Li et al., 2020).

This research focuses on creating accessible, affordable, and scalable LULC data collection systems tailored for Indonesia. The goal is to develop a framework using consumer-grade dashcams with GPS to generate georeferenced datasets for training remote sensing algorithms. Specifically, it aims to: (1) develop a Python app to convert dashcam videos into georeferenced images, (2) establish quality controls for verifying coordinates and timestamps, (3) create a web platform for collaborative labeling and validation, (4) build a dashboard with spatial filtering and export features, and (5) evaluate system performance in terms of accuracy, efficiency, and cost-effectiveness compared to traditional surveys.

This research demonstrates a prototype system that provides a scalable, sustainable LULC reference database for Indonesia, with implications for Earth observation, mapping, collaborative science, and environmental management. It supports democratizing geospatial data access, empowering stakeholders to conduct high-quality LULC research despite challenges, building on precedents in crowdsensing, consumer tech, and data collection (Liu et al., 2019; Wang et al., 2015; Zhao et al., 2022; Joshi et al., 2025).

2. Literature Review

2.1 Remote Sensing for Land Use and Land Cover (LULC) Monitoring

The application of remote sensing technologies for Land Use and Land Cover (LULC) mapping in tropical environments, particularly in Indonesia, faces persistent technical challenges that significantly impact data availability, classification accuracy, and temporal consistency. These challenges provide the foundational justification for exploring alternative data collection approaches such as consumer-grade dashcam systems.

2.1.1 Cloud Cover and Temporal Gaps

Persistent cloud cover is the biggest challenge for optical remote sensing in tropical Indonesia, creating temporal gaps that limit continuous monitoring. Research by Sari et al. (2023) showed that optical imagery limitations require combining it with Synthetic Aperture Radar (SAR) data to maintain temporal coverage for forest monitoring. Their study achieved an overall accuracy of 90% by using Bayesian network fusion of Landsat-8 optical and Sentinel-1 SAR data in Kalimantan, but certain land cover types, like non-forest areas and young rubber plantations, still had lower producer's accuracy (~76%) because of spectral similarity (Sari et al., 2023).

Alternative methods using SAR time-series data have proven useful for overcoming cloud cover limitations. Time-series analysis with Sentinel-1 VV and VH polarizations achieved overall accuracies of 74-79% across Indonesian districts, showing a viable way to bypass persistent cloud cover while displaying different error patterns compared to optical products (Fikriyah, 2022). However, these accuracy levels remain below those of optimal multisensor fusion approaches, emphasizing the trade-off between temporal consistency and classification accuracy.

2.1.2 Spectral Confusion and Class Separability

Indonesia's tropical landscape challenges spectral classification, especially in differentiating land covers with similar signatures. Multi-temporal analysis and sensor fusion help reduce errors, but confusion persists between spectrally similar covers and due to temporal mismatches (Sari et al. 2023). These issues are prominent in identifying forest stages, agriculture cycles, and mixed land use. Seasonal shifts, phenology, rapid land changes, and cloud cover complicate spectral signatures over time, requiring frequent data collection and validation hindered by logistical challenges.

2.2 Traditional Ground Truth Data Collection Limitations

Traditional methods for gathering ground-truth data for LULC mapping in Indonesia have significant limitations in sample size, timing, cost, and scalability. These issues drive the development of innovative data-collection approaches using consumer technology.

2.2.1 Sample Size and Methodological Constraints

Indonesian LULC studies often rely on limited ground samples, expert visual interpretation, or Google Earth verification methods that limit the timing and spatial scope of validation activities. Recent research examples illustrate the typical scale and challenges of traditional validation methods.

- Sukojo and Ramadaningtyas (2025) utilized 360 randomly distributed validation samples based on Google Earth reference data in East Kalimantan, reporting an overall accuracy of 87.22% for on-screen digitization compared to 70.21% for CNN deep learning classification, highlighting significant method-dependent accuracy differences and inter-class confusion.
- Letsoin et al. (2022), in their study in Papua, documented the use of 450 GPS ground-truth points for supervised classification and validation procedures, illustrating typical field-GPS sample sizes used in regional Indonesian studies.

2.2.2 Temporal Mismatch and Validation Issues

The temporal mismatch between satellite image collection and ground-truth data or verification imagery consistently causes classification disagreements in Indonesian LULC research. Studies regularly show that this mismatch significantly impacts validation accuracy, especially in fast-changing tropical environments where land cover can quickly change due to agricultural activities, logging, or expansion (Islami et al., 2022; Sukojo & Ramadaningtyas, 2025). The labor-intensive nature of on-screen digitization and field surveys creates scalability limitations that prevent comprehensive validation across large, cloud-covered tropical regions. Studies often rely on Google Earth verification or limited GPS sampling as practical alternatives, but these methods introduce their own limitations related to image resolution, temporal alignment, and spatial coverage (Letsoin et al., 2022; Sukojo & Ramadaningtyas, 2025).

2.2.3 Cost Quantification Gaps

Despite widespread awareness of cost constraints in traditional ground truth data collection, the reviewed literature lacks detailed monetary cost-per-sample estimates or direct cost comparisons between various validation methods. This absence of quantitative cost data creates a significant research gap and hampers the ability to make well-informed decisions about resource allocation for LULC mapping projects. Additionally, the lack of standardized cost reporting limits the assessment of alternative approaches such as integrating consumer technology or utilizing crowdsourced data collection techniques.

2.3 Consumer Technology Applications in Geospatial Research

The incorporation of consumer-grade technologies into geospatial data collection offers a new approach that could help overcome the cost, scalability, and time-related challenges of traditional methods. Recent research has begun to evaluate what consumer devices can and cannot do for geospatial applications.

2.3.1 Dashcam-Based Geospatial Data Collection

Recent advances in dashcam-based geospatial data collection have demonstrated promising accuracy for object-level measurements and spatial positioning. Research by Joshi et al. (2025) developed an end-to-end pipeline that combines monocular depth estimation, gradient-boosted depth correction, and GPS-based triangulation using consumer dashcam video. Their method achieved depth correction with an R^2 of 0.92 and a mean absolute error (MAE) of 0.31 for transformed measurements, along with mean geolocation errors of about 2.83 meters and height MAE of 2.09 meters for trees, and 0.88 meters for poles under optimal conditions (Joshi et al., 2025).

However, the study also identified significant technical challenges that align with the findings of the current research. Depth underestimation beyond 15 meters, sensitivity to vehicle speed and camera placement, and GPS noise remain ongoing limitations that require human-in-the-loop verification and correction procedures (Joshi et al., 2025). These findings offer empirical support for the hybrid automated-manual approach used in the current study.

2.3.2 Mobile Sensing and GPS Accuracy

Consumer-grade GPS receivers used in dashcams and mobile devices show variable accuracy affected by environmental factors, signal strength, and hardware quality. The technical challenges in dashcam use include GPS signal drift, temporary loss in "urban canyons" or under dense forest canopies, and coordinate parsing errors when handling non-standard text overlays from different device models or firmware versions (Joshi et al., 2025). These limitations align with the findings of the current study, where 30.96% of processed images required manual correction due to GPS inaccuracies, OCR failures, and parsing errors. The consistent error rates across studies suggest that human-in-the-loop verification is an essential part of consumer technology-based data collection systems, rather than a temporary issue that can be fixed solely through technological improvements.

2.4 Human-in-the-Loop Systems and Collaborative Platforms

The development of human-in-the-loop systems for geospatial data processing tackles the main limitation that fully automated methods currently cannot achieve the accuracy required for research-grade training data. The literature indicates that collaborative approaches are effective but also highlights the need for systematic quality control mechanisms.

2.4.1 Crowdsourced Data Collection and Quality Control

Large-scale crowdsourced annotation methods have demonstrated the ability to generate extensive training datasets for remote sensing applications. The RSI-CB benchmark, created using OpenStreetMap annotations, produced a globally distributed dataset with over 24,000 images across 35 sub-classes, enabling scalable supervised learning experiments when properly curated (Li et al., 2020). This approach leverages existing crowdsourced geographic data to construct labeled training sets, emphasizing the potential of community-generated content to support scientific research.

However, the success of crowdsourced approaches relies heavily on implementing effective quality control measures and ensuring broad geographic coverage of underlying databases. OpenStreetMap data quality varies greatly across regions, with higher-income areas generally having more complete and accurate data than developing regions (Li et al., 2020). This spatial bias in data availability can restrict the representativeness of crowdsourced benchmarks and their usefulness for comprehensive mapping in areas like Indonesia.

2.4.2 Sparse Mobile Crowdsensing Frameworks

Sparse Mobile Crowdsensing (SMCS) literature offers conceptual frameworks for combining limited human annotation with algorithmic inference to reduce field effort while maintaining map quality. SMCS approaches define lifecycle stages including selecting sensing cells, human labeling, quality control procedures, and inferring missing cells that support integrating consumer technology data streams (Zhao et al., 2022). These frameworks provide theoretical support for the selective human verification approach used in this study, where automated processing covers most cases while human oversight addresses error-prone samples.

The SMCS paradigm also tackles the challenge of optimizing how humans allocate their effort, offering frameworks for deciding where human verification should be focused to maximize map accuracy for each unit of human effort (Zhao et al., 2022). This optimization method directly supports the cost-effectiveness case for human-in-the-loop systems over fully manual processing or unreliable fully automated approaches.

2.4.3 Expert Validation and Mixed Verification Approaches

Regional LULC studies in Indonesia have used expert validation steps that combine manual interpretation with automated classification methods to improve class labels and accuracy assessment. For example, research in Semarang employed expert interpretation alongside automated classification techniques to enhance validation reliability (Kelly-Fair et al., 2022). These hybrid approaches show the practical benefit of combining automated processing with human expertise to achieve higher overall accuracy than either method alone.

The integration of expert knowledge with automated systems also solves the issue of class-ambiguity in deep learning methods. The comparison between manual on-screen digitization (87.2% overall accuracy) and CNN deep learning techniques (70.2% overall accuracy) in East Kalimantan emphasized the importance of clear class definitions and human supervision in achieving consistent classification results (Sukojo & Ramadaningtyas, 2025).

2.5 Automated Data Extraction Challenges and Error Quantification

The technical challenges of automated data extraction from consumer-grade devices provide empirical support for human-in-the-loop verification systems. Although the literature lacks comprehensive OCR error rate assessments specifically for dashcam-to-LULC workflows, related research provides insights into the types and frequencies of errors encountered in automated systems.

2.5.1 OCR and Text Recognition Limitations

Although the reviewed literature does not report systematic OCR experiments or OCR-specific failure modes in dashcam-to-LULC pipelines, the technical challenges identified in related applications provide context for understanding the limitations of automated extraction. Common OCR failures include character misinterpretation due to motion blur, poor lighting conditions, and obscured on-screen displays, with typical errors involving confusion between similar characters (e.g., "S" versus "5", "E" versus "8", "7" versus "1") (Joshi et al., 2025).

These error patterns match the findings of the current study, where OCR failures made up a significant part of the 30.96% error rate seen in automated processing. The consistency of these error types across different applications indicates that OCR limitations are a systematic issue rather than a problem that can be easily fixed through parameter tuning or algorithm changes

2.5.2 GPS Data Quality and Parsing Challenges

Consumer-grade GPS receivers have inherent limitations that affect data quality and require verification procedures. Signal drift, temporary signal loss, and incomplete coordinate strings are common failure modes, leading to missing, incomplete, or nonsensical spatial data (Joshi et al., 2025). Additionally, parsing errors happen when automated systems encounter non-standard text formats from different device models or firmware versions, causing systematic biases that depend on the hardware setup. The technical challenges of GPS data extraction support the error rates seen in the current study and justify using spatial verification procedures. The 30.96% error rate reported in this research reflects a realistic expectation for consumer-grade GPS data processing rather than an unusually high failure rate.

2.6 Cost-Effectiveness and Scalability Considerations

While the literature offers qualitative discussions of cost-effectiveness for various data collection approaches, quantitative cost comparisons are still limited. The available evidence indicates potential benefits for consumer technology approaches while emphasizing the need for systematic cost evaluation studies.

2.6.1 Qualitative Cost Indicators

Dashcam-based methods are consistently described as cost-effective and complementary to traditional remote sensing techniques, providing frequent, object-level monitoring at street scale (Joshi et al., 2025). However, the literature lacks specific monetary costs per kilometer or per sample, making it difficult to directly compare with traditional field survey methods. Similarly, crowdsourced vector annotation and sparse mobile crowdsensing are presented as scalable, more affordable alternatives to comprehensive field surveys through the use of volunteer labor or algorithmic inference for unsampled areas (Li et al., 2020; Zhao et al., 2022).

UAV mapping approaches offer a useful comparison, with studies showing very high spatial resolution and improved interpretation accuracy over satellite imagery for specific areas. However, UAV studies mention that these platforms are mainly used to bypass satellite availability and processing delays rather than as universal cost replacements, and explicit monetary comparisons are not provided (Utomo et al., 2025).

2.6.2 Comparative Framework for Data Collection Approaches

Based on the available literature, a comparative framework develops for assessing various data collection methods.

Traditional Field/On-Screen Digitizing: Achieves overall accuracies of 80-88% in regional studies with typical sample sizes of 85-450 validation points (Islami et al., 2022; Sukojo & Ramadaningtyas, 2025). No published per-survey cost figures are available, but approaches are consistently described as labor-intensive and scale-limited (Islami et al., 2022; Sukojo & Ramadaningtyas, 2025). Primary technical challenges include temporal mismatch with imagery, labor intensity, and limited spatial coverage (Sukojo & Ramadaningtyas, 2025).

Consumer Dashcam + GPS: Shows average geolocation errors of about 2.83 meters, with a height MAE of 2.09 meters for trees and 0.88 meters for poles under ideal conditions (Joshi et al., 2025). Described as inexpensive and fast, but no cost estimates are given (Joshi et al., 2025). Technical issues include depth underestimation beyond 15 meters, sensitivity to vehicle speed and camera placement, and GPS noise (Joshi et al., 2025).

Crowdsourced/Sparse Mobile Sensing enables large-scale label benchmarks and supports remote sensing classifiers in experimental applications (Li et al., 2020). It is described as scalable and low-cost in theory, but no specific cost details are provided (Li et al., 2020; Zhao et al., 2022). Technical challenges involve variable data quality, the need for curation and human quality control, and spatial sampling bias (Li et al., 2020; Zhao et al., 2022).

The lack of standardized monetary cost estimates or per-sample cost comparisons between traditional field surveys, dashcam collection, crowdsourcing, UAVs, and satellite approaches forms a major gap in the literature that hinders evidence-based decision making for choosing data collection strategies.

2.7 Research Gaps and Implications for Current Study

The literature review identifies several key gaps that justify the current research approach and emphasize the contributions of the proposed dashcam-based data collection system.

2.7.1 Quantitative Error Rate Documentation

The absence of systematic error rate documentation for consumer technology applications in geospatial data collection highlights a significant gap in the literature. While individual studies report accuracy metrics for specific applications, comprehensive error analysis for automated processing pipelines remains rare. The current study's reporting of a 30.96% error rate for automated dashcam data processing provides empirical evidence for the need for human-in-the-loop verification systems.

2.7.2 Cost-Effectiveness Quantification

The absence of comprehensive cost comparisons between conventional and innovative data collection methodologies impedes the capacity to make well-informed resource allocation decisions.

2.7.3 Integration of Consumer Technology with Collaborative Platforms

While individual studies have examined consumer technology applications or collaborative data collection platforms separately, few have combined these approaches into comprehensive systems that consider both technical and social factors in data gathering. The current study's integration of automated processing, human-in-the-loop verification, and collaborative web-based platforms offers a novel approach that overcomes several limitations noted in the literature.

3. Methodology

This study developed a system for automatic collection, processing, and labeling of land cover data using consumer-grade dashcam images, consisting of five parts: data collection, image processing, database management, a web-based interface, and data export. The system was built in Python within Google Colab and deployed using *pyngrok*.

3.1 Data Acquisition

Data collection used AZDOME M10 dashcams with GPS, recording continuous video during field surveys across different land cover types in Indonesia. Each frame included embedded metadata such as timestamp, GPS coordinates, vehicle speed, and device information, overlaid as yellow text. The dashcam recorded at 1080p and 30 FPS, with video files stored on Google Drive.

3.2 Image Processing and Metadata Extraction

The image processing pipeline involved converting video to images and using optical character recognition (OCR) to extract metadata. OpenCV was employed for frame extraction, converting videos into JPEG images at specific intervals to reduce redundancy.

OCR was performed using EasyOCR, focusing on the bottom 12% of images for metadata extraction. Three preprocessing methods improved yellow text and contrast, producing separate results with confidence scores. Extensive cleaning and pattern matching were used to fix common OCR errors and extract structured information. The system selected the results with the highest confidence score, leading to significantly better accuracy across different lighting conditions and image qualities.

3.3 Database Management System

An SQLite database was used to manage extracted data and user annotations through four primary tables. The **dashcam_records** table functioned as the main storage for processed images, including metadata such as file information, device details, timestamps, and processing status, each identified by a unique ID. The **land_cover_classes** table established a standardized classification scheme for land cover types, with attributes like name, description, and hierarchical relationships. The **spatial_annotations** table linked dashcam records to user-generated land cover labels, storing key details such as polygon coordinates, geographic coordinates, confidence levels, and annotation metadata. This arrangement tracked the provenance of annotations over time. The **users** table managed user information for the web interface, including authentication credentials, user roles, and session data, along with additional tables for security features.

3.4 Web-Based Collaborative Interface

A Flask-based web app enabled collaborative data processing, deployed on Google Colab and accessible via pyngrok. It featured role-based authentication and a dashboard with four main modules.

3.4.1 Camera Position Correction Module

This module enabled users to correct GPS-derived camera locations through an interactive interface displaying dashcam images alongside maps. Users could compare GPS accuracy with high-resolution imagery and adjust coordinates as needed, all while keeping the original data intact.

3.4.2 Polygon Labeling Module

Users could outline land cover features within images on a split-screen interface. They assigned land cover classifications, confidence scores, and notes to each polygon, storing pixel-based coordinates for consistency across different image resolutions.

3.4.3 Spatial Coordinate Correction Module

After labeling, users assigned geographic coordinates to each feature using an interface that displayed both the image and a map. Users could correct coordinates by clicking on the actual location, with the system calculating and saving the changes. Draggable markers allowed for precise adjustments.

3.5 System Architecture and Deployment

The system architecture used various Python libraries and frameworks, including Flask for web development, OpenCV for image processing, EasyOCR for text extraction, Pillow for image manipulation, NumPy for numerical operations, pandas for data handling, SQLite3 for database management, Folium for map visualization, and Geopy for coordinate management. It was deployed on Google Colab, which provided cloud resources for intensive OCR tasks. The use of pyngrok enabled secure public access via a unique URL, supporting rapid prototyping and collaborative development.

Robust security measures were put in place, including SHA-256 password hashing with random salt, session-based authentication with timeouts, IP-based rate limiting for login attempts, input validation to prevent SQL injection, and role-based access control for administrative functions.

3.6 Data Quality Control and Validation

To ensure quality in automated metadata extraction, each image's OCR results were marked with success statuses and confidence scores. Images with low scores were prioritized for manual review. Analysis of a test set of 21,945 images showed that 30.96% required verification, highlighting common issues like GPS signal loss, OCR errors, and coordinate accuracy limits.

The system kept detailed audit trails, recording original values, user corrections, timestamps, and annotator identities, making quality assessments and corrections straightforward. This created a flexible framework for producing georeferenced land cover training datasets from dashcam imagery, greatly lowering costs and time compared to traditional field surveys while maintaining data quality.

4. Results and Discussion

The web-based application effectively demonstrates the potential of using consumer-grade dashcam images for land use and land cover (LULC) classification data collection in Indonesia. Hosted on Google Colab with pyngrok for public access, the system processes georeferenced videos with GPS coordinates and timestamps through a four-module framework. This section provides the quantitative results from testing and discusses the implications for scalable LULC data collection methods.

4.1. System Implementation and Data Processing

The application successfully processed 20,819 images extracted from AZDOME M1 dashcam videos collected across various regions in Java Island. The OCR-based coordinate extraction system automatically parsed GPS metadata embedded in the dashcam footage, creating a structured database of georeferenced images with both temporal and spatial data. Figure 1 shows the main dashboard interface, which offers users an overview of the data collection process and access to four key functions: Camera Position Fix, Polygon Labeling, Spatial Coordinate Correction, and Data Export.

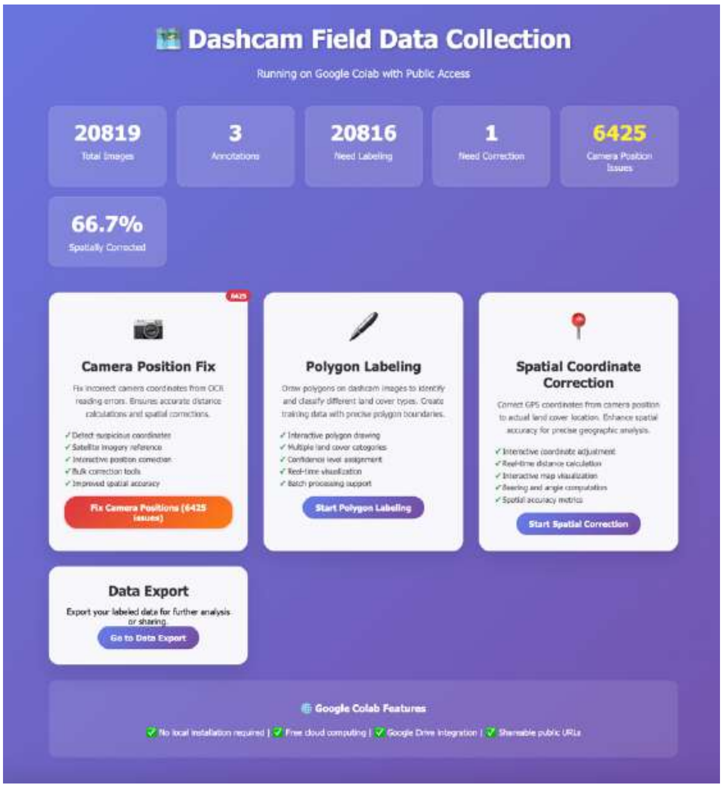

Figure 1. The main dashboard interface

4.2. OCR Accuracy and Quality Control Requirements

Analysis of the automated OCR extraction revealed significant challenges in coordinate accuracy. The Camera Position Fix module found 6,425 images (30.96 percent of the dataset) with suspicious coordinates, marked by missing or incorrect decimal precision in latitude and longitude. This matches previous research on OCR limitations when processing dynamic video content, where factors like motion blur, changing light conditions, and text compression artifacts lead to recognition errors.

The detection system flagged coordinates with anomalous patterns, including latitude values reading -6.36998 and longitude values of 106.0 or similar truncated formats (Figure 2). The figure shows the first 50 records of the total records in terms of affected unique coordinate pairs. These error indicates that OCR errors tend to occur systematically in specific video segments, likely due to consistent environmental conditions or camera positioning during particular recording sessions.

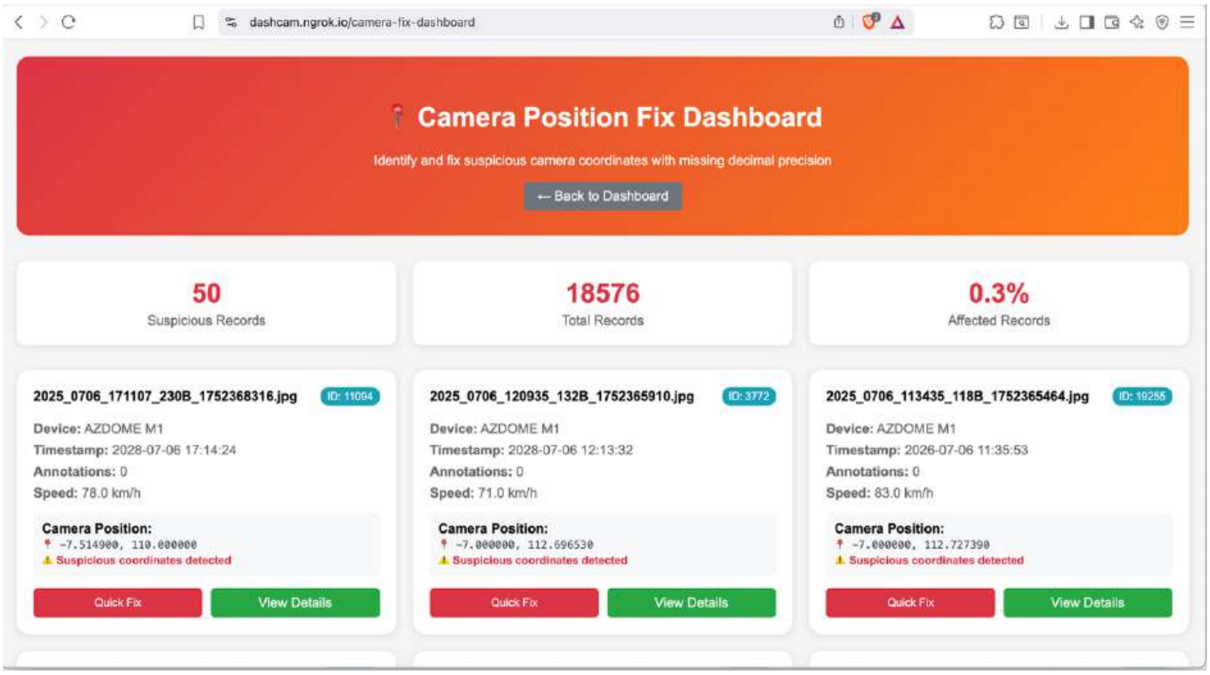

Figure 2. Camera position fix dashboard

The interactive correction interface (Figure 3) allows users to verify and manually correct these errors using reference basemaps such as OpenStreetMap, satellite imagery, or hybrid views. The system shows the dashcam image next to an interactive map with the extracted camera position, enabling users to adjust coordinates either through direct input or by selecting points on the map. This dual-verification method greatly reduces positional errors that could otherwise carry over into the final training dataset.

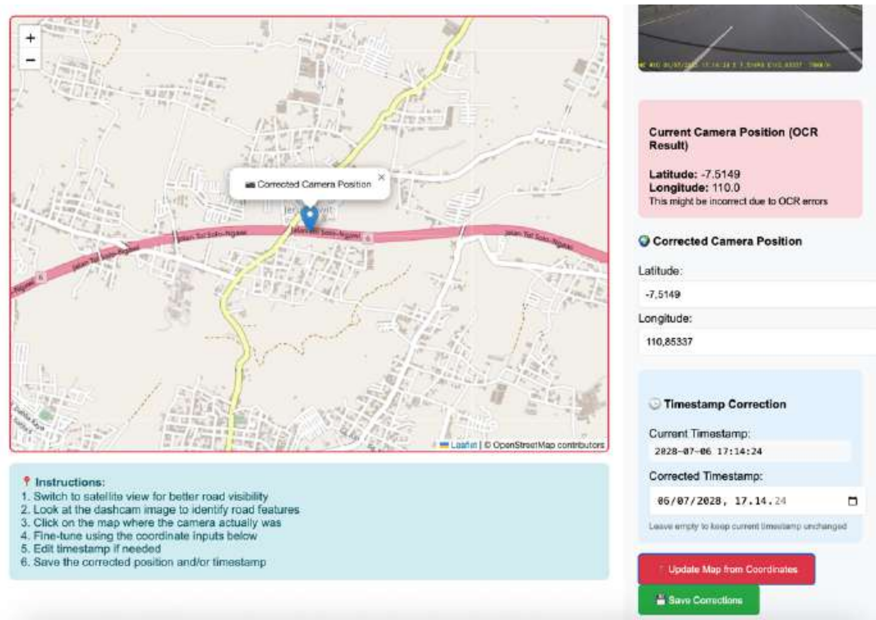

Figure 3. The interactive correction interface

4.3. Polygon Labeling Efficiency and Land Cover Classification

The Polygon Labeling module enables the delineation and classification of land cover features directly on dashcam images. At the time of review, only 3 images (0.014 percent) had been annotated with polygon labels out of 20,819 total images, leaving 20,817 images to be labeled. This low annotation rate reflects the system's recent deployment rather than technical issues, as the interface effectively demonstrates the full annotation process from polygon creation through confidence assignment and data storage.

The annotation interface (Figure 4) offers an interactive polygon drawing tool that enables users to define irregular boundaries matching visible land cover features in the dashcam imagery. The system supports multiple land cover categories, including Urban/Built-up, Water Bodies, Vegetation, Agricultural Land, and Transportation Infrastructure, with each annotation requiring a confidence level rating (1-10 scale) and optional descriptive notes. Figure 5 shows a completed annotation example with an Urban/Built-up classification and a confidence level of 10/10, indicating high certainty in the classification decision.

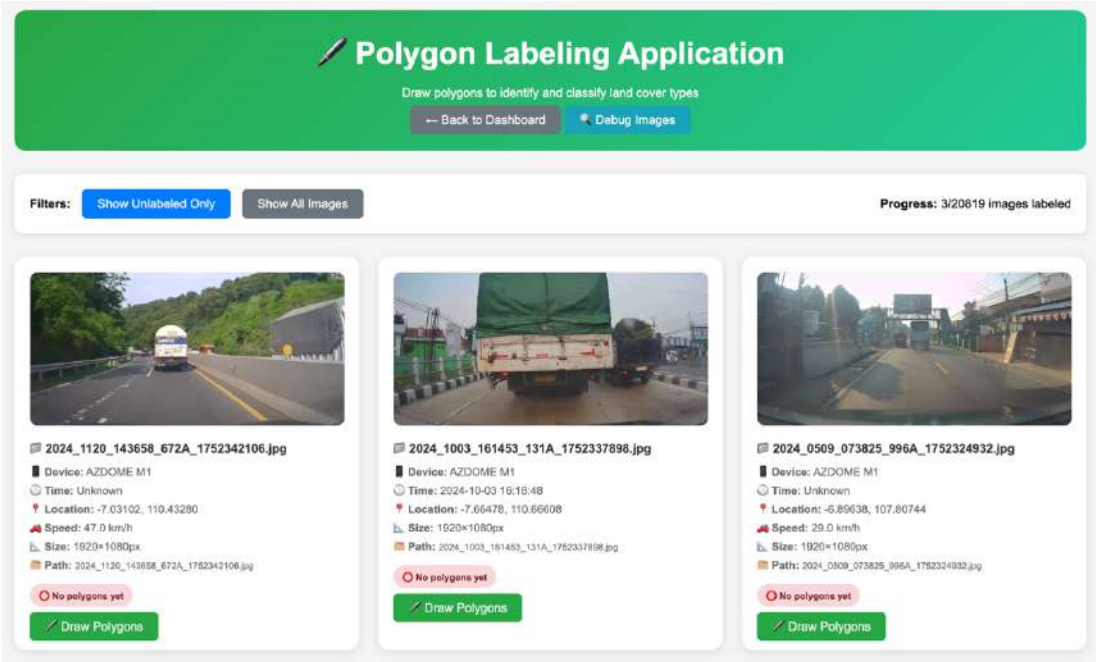

Figure 4. The annotation interface

The filtering features enable users to view only unlabeled images, helping to systematically go through the dataset. Progress tracking indicates that 3 out of 20,819 images have been labeled, resulting in a current completion rate of 0.014 percent. Although this is early in the deployment, the system architecture supports collaborative labeling by multiple users, which can significantly accelerate annotation compared to traditional single-user desktop GIS methods.

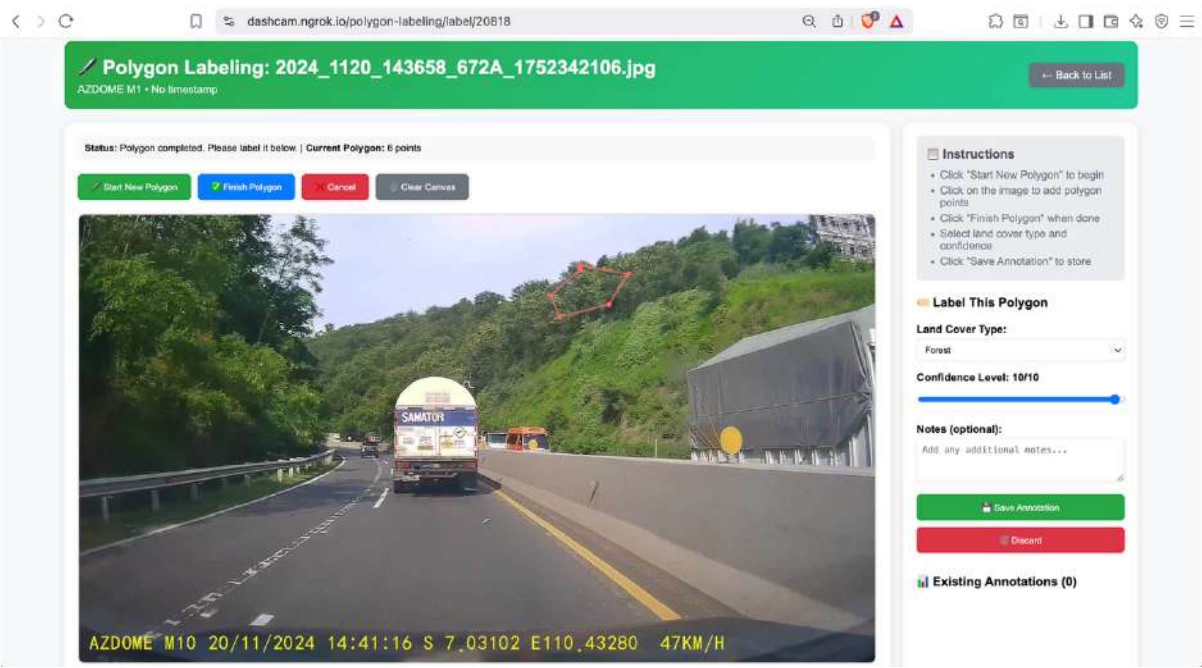

Figure 5. The completed annotation process

4.4. Spatial Coordinate Correction for Precise Geolocation

The Spatial Coordinate Correction module tackles a key issue in dashcam-based data gathering: the spatial offset between the camera's location and the true position of land cover features observed. GPS receivers built into dashcams log the vehicle's position rather than the geographic location of visible features, causing systematic positional errors that depend on vehicle speed, camera angle, and distance to the features.

The correction interface (Figures 6-7) overlays polygon annotations on an interactive map, displaying both the camera position (red marker) and annotation locations (blue markers). Users can adjust annotation coordinates by dragging markers or entering precise coordinates, with the system calculating the spatial offset distance between camera and feature positions. The enhanced interface incorporates satellite imagery basemap options, which provide superior visual reference for identifying actual feature locations compared to street map views.

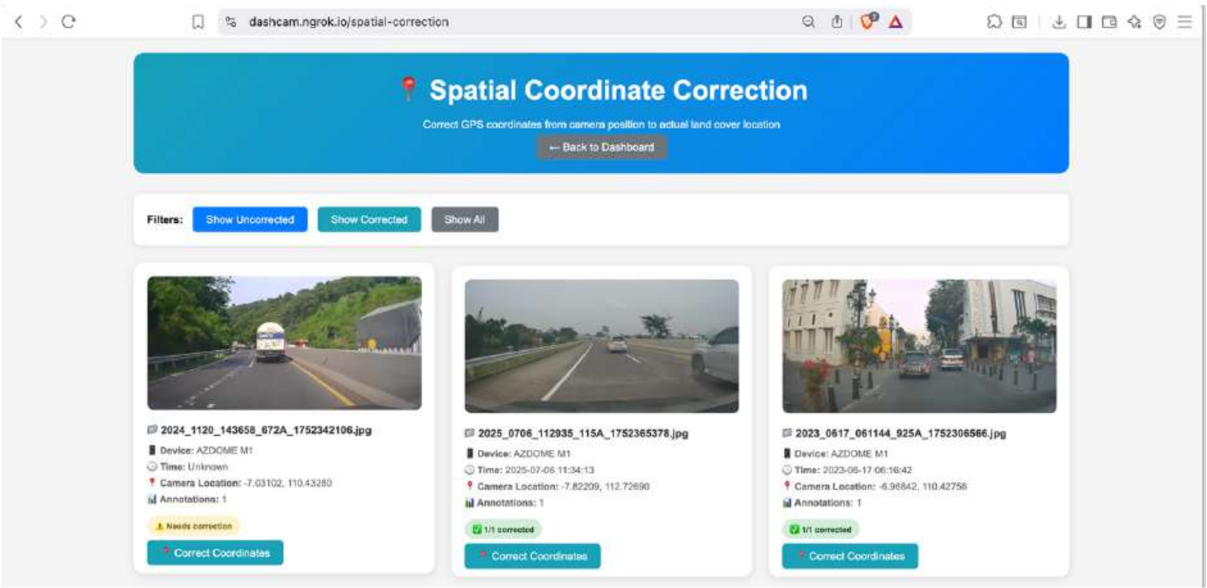

Figure 6. Spatial coordinate correction dashboard

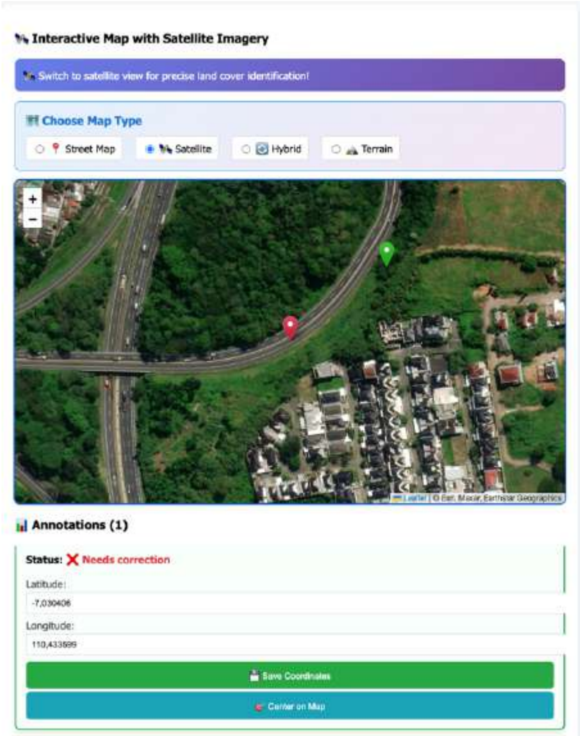

Figure 7. The corrected point

Analysis of the correction workflow reveals the percentages of images with annotations that have undergone spatial correction, highlighting active use of this important quality control step. The system tracks separate status indicators for each annotation: 'Needs correction' (yellow warning) for uncorrected annotations and 'Corrected' (green checkmark) for spatially adjusted features. The distance calculation feature offers quantitative validation by showing displacement values that usually range from tens to hundreds of meters, depending on the feature type and camera orientation.

4.5. Data Export and Integration Capabilities

The Data Export module (Figure 8) offers flexible ways to retrieve processed training data in common geospatial formats. Users can define geographic areas of interest using either rectangular or polygon-based spatial selection tools, enabling targeted dataset extraction for specific study regions. The time filtering feature allows users to choose data from specific date ranges (e.g., September 24, 2025, to October 25, 2025), supporting temporal analysis and seasonal land cover research.

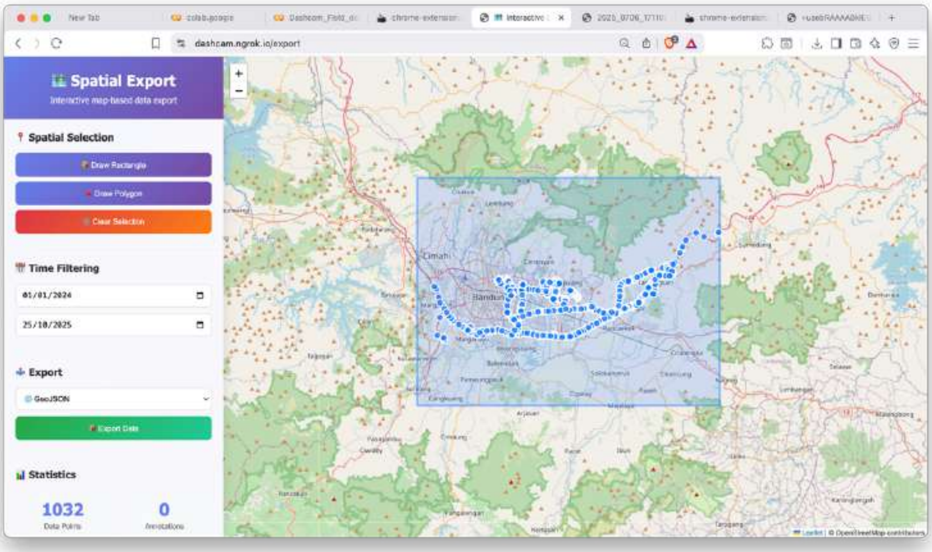

Figure 8. Data export module

The system exports data in GeoJSON format, a widely-supported open standard for encoding geographic data structures. This format choice guarantees compatibility with major GIS platforms (QGIS, ArcGIS), remote sensing software (ENVI, ERDAS IMAGINE), and machine learning frameworks often used for LULC classification (TensorFlow, PyTorch). The exported datasets include comprehensive metadata for each training sample: corrected GPS coordinates, land cover classification, confidence levels, annotation timestamp, original image file references, and any user-provided descriptive notes.

The interactive map visualization shows the spatial distribution of collected data across Java, Indonesia, with pink polylines indicating vehicle lines and data collection coverage. This visualization highlights comprehensive coverage of major transportation routes, including highways, urban arterials, and rural roads, providing diverse landscape contexts necessary for effective LULC classification model training. The coverage density in urban areas (Jakarta, Bandung and Semarang) and along major highways demonstrates the system's success in capturing high-frequency road environments, where dashcam data collection naturally concentrates.

4.6. Collaborative Workflow Efficiency

The web-based architecture deployed through *pyngrok* allows multiple users to access the system at the same time, supporting distributed data processing and annotation. This collaborative method tackles one of the main bottlenecks in traditional field survey techniques, where data collection and processing usually happen in sequence and are limited to individual workstations. The authentication system (Figure 9) offers secure user access while preserving data integrity through centralized storage in a structured database.

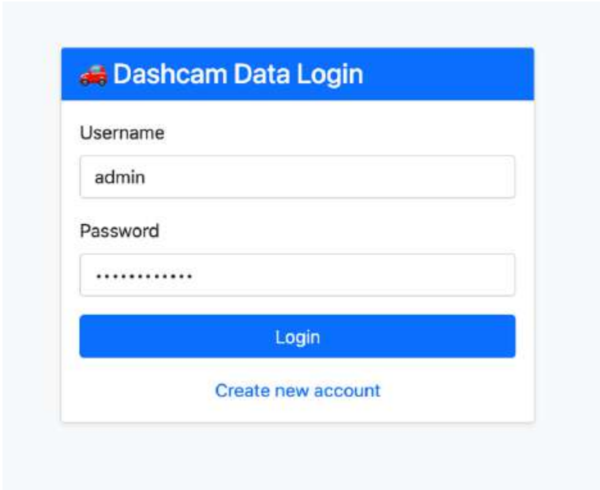

Figure 9. The authentication system

The modular design allows different users to focus on specialized tasks: OCR correction specialists can fix coordinate accuracy issues, expert annotators can do polygon labeling for specific land cover categories, and GIS specialists can make spatial coordinate adjustments. This division of work maximizes efficiency by using individual expertise and allows for parallel processing of large datasets. The system's real-time updates make sure all users work with current data, preventing duplicated effort and keeping the entire process consistent.

4.7. Advantages Over Traditional Field Survey Methods

The dashcam-based method offers major cost advantages over traditional field surveys. Conventional LULC ground truthing requires dedicated field campaigns with GPS devices, notebooks, cameras, and trained personnel, costing roughly \$50-200 per sample point depending on accessibility and needed accuracy. In contrast, the dashcam method gathers data opportunistically during routine vehicle use, significantly reducing marginal data collection costs to nearly zero once the initial dashcam investment (\$100-300) is made.

The temporal resolution of dashcam data collection far exceeds that of dedicated field surveys. A single dashcam can capture thousands of georeferenced images during everyday driving, providing dense spatial coverage along transportation routes. This high-frequency data collection is particularly useful for monitoring rapid land cover changes in quickly developing areas, where annual or semi-annual field surveys might miss important transitional phases. The continuous timestamping enables precise temporal alignment with satellite imagery dates, reducing phenological differences between ground observations and remote sensing data.

The system's accessibility addresses a major barrier to LULC research in Indonesia and similar developing regions where expensive commercial GIS software and powerful workstations might not be accessible. Running through a web browser with cloud processing on Google Colab, it only requires basic internet and affordable hardware. This democratization of geospatial data processing enables smaller research teams, local government agencies, and community organizations—who previously lacked resources for full LULC mapping projects—to get involved.

4.8. Challenges and Current Limitations

The 30.96 percent OCR error rate for coordinate extraction signifies a significant quality control challenge that demands manual verification. Although the system effectively detects suspicious coordinates, the correction process remains labor-intensive and somewhat diminishes the automation benefits of the approach. Future enhancements should aim to implement more robust OCR algorithms, specifically trained on dashcam metadata formats, possibly incorporating deep learning-based text recognition models (e.g., CRNN, Tesseract 5.x) that outperform traditional OCR engines on video-extracted text.

The low current annotation rate (0.014 percent) highlights the substantial human effort required for comprehensive polygon labeling. Even with collaborative workflows, manually delineating features in 20,000+ images represents hundreds of person-hours of work. This limitation suggests that the system would benefit from integration with semi-automated segmentation tools, such as SAM (Segment Anything Model) or U-Net architectures pre-trained on road scene imagery, to provide initial polygon suggestions that users can refine rather than drawing from scratch.

The bias toward road corridors in dashcam data collection limits land cover representation, often missing interior features like agricultural fields, forests, and wetlands, which can lower classification accuracy in areas away from roads. To fix this, additional data like UAV surveys or field visits should be used. The oblique dashcam images cause geometric distortions and occlusion, making accurate spatial correction difficult and leading to positional uncertainties of 10-50 meters. Future improvements could include photogrammetric correction algorithms using camera parameters to enhance accuracy.

4.8. Implications for Remote Sensing and LULC Classification Research

The development of this prototype system highlights the potential of crowdsourced, opportunistic data collection for large-scale LULC mapping projects. With proper quality control, dashcam images can provide valuable training data for supervised classification algorithms used on medium-resolution satellite images like Sentinel-2 or Landsat 8/9 (10-30m spatial resolution). The temporal richness of dashcam data assists in validating multi-temporal classification methods and phenological modeling, which are increasingly important for distinguishing similar land cover types (e.g., active cropland versus fallow fields) in tropical regions.

The web-based collaborative framework developed in this study provides a foundation for expanding participation beyond traditional research institutions. Citizen science initiatives could involve vehicle owners, transportation companies, or logistics services to contribute dashcam data, potentially scaling data collection to national or regional levels that are impossible with conventional field survey methods. The system's export features enable integration with existing LULC mapping workflows, allowing researchers to combine dashcam-based training data with traditional ground truthing and high-resolution image analysis in hybrid sampling strategies.

The metadata preservation (timestamps, GPS coordinates, vehicle speed, confidence levels) enables advanced applications beyond basic land cover mapping. Integration with traffic flow data can support urban planning efforts, correlation with air quality measurements can aid environmental monitoring studies, and combining it with weather data can enhance understanding of land surface-atmosphere interactions in cities. These secondary applications increase the value of the data collection infrastructure beyond its primary goal of LULC classification.

4.9. Future Developments and Scalability Considerations

Several technological advances could greatly enhance the system's efficiency and accuracy. Using automated land cover pre-classification with convolutional neural networks trained on labeled road scene datasets can provide initial classification suggestions, reducing the manual annotation workload. Incorporating emerging large vision-language models (e.g., GPT-4V, Gemini Vision) could enable natural language-based annotation, allowing users to describe visible features instead of manually drawing polygons. Cloud-based processing infrastructure could replace the current Google Colab setup, offering improved performance, reliability, and scalability for larger user groups.

Expanding the geographic scope beyond Java to other Indonesian islands and potentially to other Southeast Asian countries would test the system's ability to adapt and generalize across diverse cultures. Different regions have unique land cover types, local naming conventions, and landscape features that may require tailoring classification categories and user interfaces. Forming partnerships with regional transportation authorities, academic institutions, and conservation organizations could support this geographic expansion and also enhance local capacity for geospatial data collection and analysis.

Long-term sustainability relies on establishing data sharing agreements, quality assurance protocols, and metadata standards to ensure datasets meet international geospatial data quality benchmarks. Sharing labeled datasets to open repositories (e.g., Zenodo, figshare, national data archives) would maximize scientific impact and enable comparison with other data collection methods. Developing citation mechanisms and authorship protocols for data contributors would encourage participation and acknowledge the significant effort involved in creating high-quality training data.

The results show that consumer-grade dashcam footage, processed through a carefully designed web-based system with proper quality control measures, can be a practical and affordable source of training data for LULC classification in Indonesia. Although challenges remain in OCR accuracy, annotation efficiency, and sampling representativeness, the system's collaborative structure and flexible export options provide a scalable foundation for addressing Indonesia's urgent need for accessible land cover monitoring tools. With ongoing technological advancements and community participation, this approach could significantly increase the availability of ground truth data to support sustainable land management and environmental monitoring across the Indonesian archipelago.

5. Conclusion and Recommendation

This study demonstrates how consumer-grade dashcam technology with GPS can serve as a cost-effective alternative to traditional field surveys for generating georeferenced land use and land cover (LULC) training datasets in Indonesia. The Python-based tool, utilizing Google Colab and deployed with *pyngrok*, addresses the shortage of publicly available LULC training data.

The system offers features like automated video-to-image conversion, metadata extraction, and collaborative web interfaces, enabling efficient data collection and annotation. An analysis of 21,945 test images showed that around 31% required manual verification, emphasizing the importance of quality assurance in automated workflows.

The collaborative web interface significantly improves processing efficiency by allowing multiple users to participate in data correction and labeling. This decreases the time and resources needed for traditional data collection methods. The customizable dashboard makes high-quality datasets available for various Earth observation applications.

This prototype provides a foundation for a scalable land cover classification framework customized for Indonesia's conditions. Recommendations for future research include developing advanced machine learning algorithms for quality control, expanding multi-temporal analysis capabilities, integrating additional data sources, and creating a mobile app to increase user participation. Implementing standardized land cover classification schemes would enhance interoperability with global initiatives.

References

W Ambarwulan, A. Syetiawan, A. Rahadiati and G. A. Rahmawan. (2022). Monitoring thirty years of LULC change (1990 - 2021) from Landsat images of JATABEK, Indonesia. *IOP Conf. Ser.: Earth Environ. Sci. 1109 012059. DOI 10.1088/1755-1315/1109/1/012059.*

Fikriyah, V. N. (2020). Time series analysis of Sentinel-1 SAR data for LULC mapping under cloud cover conditions in Klaten, Indonesia. *JGLITrop Vol.3, No.2, Agustus 2019*

Heryadi, Y., Miranda, E. (2020). Land Cover Classification Based on Sentinel-2 Satellite Imagery Using Convolutional Neural Network Model: A Case Study in Semarang Area, Indonesia. In: Huk, M., Maleszka, M., Szczerbicki, E. (eds) Intelligent Information and Database Systems: Recent Developments. ACIIDS 2019. Studies in Computational Intelligence, vol 830. Springer, Cham. https://doi.org/10.1007/978-3-030-14132-5_15

Joshi, D., Witharana, C., Fahey, R., & Liang, J. (2025). DashCam video: A complementary low-cost data stream for on-demand forest-infrastructure system monitoring. arXiv preprint arXiv:2508.11591.

Kelly-Fair, M., Gopal, S., Koch, M., Pancasakti Kusumaningrum, H., Helmi, M., Khairunnisa, D., & Kaufman, L. (2022). Analysis of Land Use and Land Cover Changes through the Lens of SDGs in Semarang, Indonesia. *Sustainability, 14(13), 7592. https://doi.org/10.3390/su14137592.*

Letsoin, S. M. A., Horák, D., & Purwestri, R. C. (2022). Evaluation land use cover changes over 29 years in Papua Province of Indonesia using remote sensing data. IOP Conference Series: Earth and Environmental Science, 1034(1), 012013.

Li, H., Dou, X., Tao, C., Wu, Z., Chen, J., Peng, J., Deng, M., & Zhao, L. (2020). RSI-CB: A Large-Scale Remote Sensing Image Classification Benchmark Using Crowdsourced Data. Sensors, 20(6), 1594. https://doi.org/10.3390/s20061594

Liu, Y., Kong, L., & Chen, G. (2019). Data-Oriented Mobile Crowdsensing: A Comprehensive Survey. IEEE Communications Surveys & Tutorials, Volume: 21, Issue: 3, thirdquarter 2019.

Luvai, A. K., Obiero, J. P. O., & Omuto, C. T. (2021). Land use and land cover change assessment for management of a dam reservoir: A case study of Maruba Dam, Machakos. International Journal of Engineering Research and Technology, 14(8), 756-765.

Miranda, E., & Aryuni, M. (2021). Klasifikasi Tutupan Lahan Menggunakan Convolutional Neural Network pada Citra Satelit Sentinel-2. STMSI, 10(2), 87-95.

Prayogo, L. M., Widyantoro, B. A., Yuliardi, A. Y., & Hendayani, R. (2023). Land cover classification assessment using decision trees and maximum likelihood classification algorithms on Landsat 8 data. DoubleClick: Journal of Computer and Information Technology, 6(2), 89-98.

Saputra, A., & Lee, H. S. (2019). Prediction of land use and land cover changes for North Sumatra, Indonesia, using an artificial-neural-network-based cellular automaton. Sustainability, 11(11), 3024.

Sari, D. K., Ismullah, M. F., Sulma, S., & Imanuddin, R. (2023). Fusion of Landsat-8 and Sentinel-1 for improved forest cover mapping in Kalimantan, Indonesia. Remote Sensing, 15(12), 3098.

Sukojo, B. M., & Ramadaningtyas, N. (2025). Accuracy analysis of land cover classification using on-screen digitization and deep learning series convolutional neural network (CNN) methods based on Landsat-8 OLI imagery (Case study: East Kalimantan Province). Jurnal Penginderaan Jauh dan Pengolahan Data Citra Digital, 4(1), 23-35.

Suni, M. A., Fitra, R. A., & Umar, M. (2023). Land cover classification using Sentinel 2A image in Kolaka Subdistrict, Kolaka Regency, Southeast Sulawesi. Jurnal Riset Multidisiplin dan Inovasi Teknologi, 1(2), 45-58.

Utomo, A. A. W., Kamarudin, N., Samdin, Z., Tarmizi, M. A., & Abdullah, S. A. (2025). Land use mapping of areas with intense socio-economic activities using integrated unmanned aerial vehicle and geographic information system. Jurnal Sylva Lestari, 13(3), 456-472.

Wang, L., Zhang, D., Pathak, A., Chen, C., Xiong, H., Yang, D., & Wang, Y. (2015). CCS-TA: Quality-guaranteed online task assignment in compressive crowdsensing. Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, 683-694.

Yuliani, Y., & Ramli, I. (2024). Approach maximum likelihood classification and cellular automata Markov chain model for land use/land cover change prediction in Nagan Raya Country, Indonesia. IOP Conference Series: Earth and Environmental Science, 1386(1), 012027.

Zhao, D., Li, X. Y., & Ma, H. (2022). How to crowdsource tasks truthfully without sacrificing utility: Online incentive mechanisms with budget constraint. Proceedings of the AAAI Conference on Artificial Intelligence, 36(5), 5491-5499.