THESIS PROPOSAL

Title: Leveraging Siamese U-Net with Attention Mechanisms for Enhanced Deforestation Risk Assessment and Smallholder Coffee Plot Delineation in Indonesia using Sentinel-2 Imagery for EUDR Compliance

Abstract:
The European Union Deforestation Regulation (EUDR) imposes stringent traceability and deforestation-free requirements on commodities like coffee, significantly impacting Indonesian producers, especially smallholders. Artificial Intelligence (AI) offers promising solutions for compliance, particularly through remote sensing for deforestation monitoring and plot delineation. This thesis aims to design, develop, and evaluate a Siamese U-Net architecture incorporating attention mechanisms for accurate deforestation risk assessment and precise delineation of smallholder coffee plots in selected regions of Indonesia, utilizing Sentinel-2 satellite imagery. The proposed methodology involves: (1) Curating a dataset of Sentinel-2 imagery and corresponding ground-truth data for coffee plots and deforestation events in Indonesia, leveraging existing datasets and potentially limited primary data collection. (2) Pre-processing satellite imagery, including atmospheric correction and cloud masking suitable for tropical regions. (3) Designing and training a Siamese U-Net with attention for change detection (deforestation) and semantic segmentation (plot delineation). (4) Evaluating model performance using metrics such as F1-score, Intersection over Union (IoU), and overall accuracy, and comparing it against a baseline U-Net. The research is expected to deliver a robust AI model capable of identifying deforestation linked to coffee production and delineating smallholder plots with improved accuracy over standard approaches. This will demonstrate a feasible AI-driven tool to support Indonesian stakeholders in EUDR due diligence, contributing to sustainable coffee supply chains and informing national monitoring efforts.

1. Introduction

1.1. Detail of Background of Research

The global effort to combat deforestation has gained significant momentum with the introduction of stringent regulatory measures by major consumer markets. Among these, the European Union Deforestation Regulation (EUDR) stands out as a transformative piece of legislation with far-reaching implications for international trade and environmental governance.

- The European Union Deforestation Regulation (EUDR): A Paradigm Shift in Global Trade
  The European Union Deforestation Regulation (EUDR), formally Regulation (EU) 2023/1115, establishes a comprehensive framework aimed at ensuring that products consumed within the EU do not contribute to global deforestation or

forest degradation.1 It prohibits the placing on the EU market, or export therefrom, of seven key commodities—cattle, cocoa, coffee, oil palm, rubber, soya, and wood—and a wide range of derived products, unless they meet three critical conditions: they must be 'deforestation-free', produced in accordance with the relevant legislation of the country of origin, and covered by a...source 2] The term 'deforestation-free' stipulates that the commodities must have been produced on land that was not subject to deforestation after December 31, 2020. For wood products, an additional requirement is that they must have been harvested without inducing 'forest degradation' after the same cut-off date.2 Forest degradation, in this context, refers to the conversion of primary forests or naturally regenerating forests into plantation forests or other wooded land.2 This regulation is a cornerstone of the EU's Green Deal, reflecting a commitment to minimizing the EU's environmental impact on a global scale.1

A central pillar of the EUDR is the mandatory due diligence obligation imposed on operators (companies placing the products on the EU market for the first time or exporting them).[1] This process comprises three distinct steps. Firstly, **Information Collection**, where operators must gather comprehensive data, including the precise geolocation coordinates (latitude and longitude) of all plots of land where the relevant commodities were produced, the quantity of commodities, country of production, and verifiable information that the commodities are deforestation-free and legally produced.[2] Secondly, **Risk Assessment**, where operators must analyze the collected information to assess the risk of non-compliance. This assessment considers factors such as the prevalence of deforestation in the country or region of production, the complexity of the supply chain, the risk of commodities being mixed with those of unknown origin, and information provided by suppliers.[2] The EU will also establish a country benchmarking system, classifying countries as low, standard, or high risk, which will influence the stringency of due diligence, although all countries are considered standard risk until at least June 2025.[2] Thirdly, **Risk Mitigation**, whereby if the assessment reveals a non-negligible risk, operators must adopt adequate and proportionate mitigation measures to reduce that risk to a negligible level. Such measures may include requiring additional information, conducting independent surveys or audits, or supporting suppliers, particularly smallholders, in achieving compliance through capacity building and investments.[2] Before placing products on the market, operators must submit a due diligence statement through a centralized EU information system, known as TRACES.[2]

Failure to comply with the EUDR can lead to severe penalties. These include fines of up to at least 4% of the operator's total annual Union-wide turnover in the preceding financial year, temporary exclusion from public procurement processes, temporary prohibition from placing or making available on the market or exporting the relevant commodities and products, and confiscation of the relevant products or revenues gained.[1]

The EUDR entered into force on June 29, 2023. The new rules will become applicable to medium and large operators from **December 30, 2025**, and to micro and small enterprises (SMEs) from **June 30, 2026**.[1] This timeline reflects a 12-month postponement from the initially planned dates, granted to allow stakeholders more time for preparation.[1]

The EUDR's design inherently links environmental objectives with market access mechanisms. The regulation's primary aim is to curb deforestation and forest degradation driven by EU consumption and production.[1] This environmental protection goal is enforced by prohibiting non-compliant products from entering or being exported from the EU market.[2] Consequently, for major commodity-producing nations such as Indonesia, adherence to the EUDR is not merely an environmental consideration but a critical economic imperative to maintain trade relations with one of its significant export markets.[5] This dual mandate of environmental safeguarding and market regulation intensifies the pressure on producer countries to adopt robust, verifiable systems and technologies to demonstrate compliance, as failure to do so carries direct and substantial economic consequences, including loss of market access and financial penalties.[1] This heightened pressure underscores the urgency for effective solutions, including those leveraging Artificial Intelligence.

- Indonesia's Vulnerability and Response to EUDR

  Indonesia, as one of the world's foremost producers and exporters of several commodities covered by the EUDR—notably palm oil (being the largest global producer and a major supplier to the EU 16), timber, coffee, cocoa, and rubber—finds itself significantly affected by this regulation.5 The EUDR's requirements pose substantial challenges to the Indonesian agricultural sector and its export economy.

  A primary area of concern revolves around the **smallholder farmers**, who constitute the backbone of production for many of these commodities, particularly palm oil and coffee.[5] These smallholders, often operating with limited financial, technical, and informational resources, face considerable difficulties in meeting the EUDR's complex demands. These include establishing detailed traceability systems, conducting GPS mapping of their often small and irregularly shaped plots, obtaining formal land tenure documentation such as the Surat Tanda Daftar Budidaya (STDB), and navigating intricate international regulatory frameworks.[7] The administrative burden associated with these requirements is

disproportionately high for small-scale producers.[7]
The EUDR's mandate for **traceability and geolocation**, requiring commodities to be traced back to the precise plot of land of production using geographic coordinates, presents a major hurdle.[4] This is particularly challenging within Indonesia's often fragmented and multi-layered supply chains, which involve numerous smallholders and intermediaries before products reach export channels.[16] Furthermore, **data discrepancies and definitional ambiguities**, such as differing interpretations of "forest" and inconsistencies between Indonesian national forest maps and data references used by the EU, have been sources of concern and potential compliance friction.[7] The requirement for **legality verification** also adds complexity, as operators must demonstrate that products were produced in accordance with a wide range of national laws in the country of origin. These laws pertain to land use rights, environmental protection, forest management, biodiversity conservation, the rights of third parties (including the Free, Prior, and Informed Consent - FPIC of Indigenous Peoples and local communities), labor rights, human rights protected under international law, and regulations concerning tax, anti-corruption, trade, and customs.[2]
Recognizing these challenges, the Indonesian government has initiated several proactive measures. A joint task force (Satgas EUDR) has been established with Malaysia to represent the interests of palm oil smallholders and engage with the EU.[16] A significant national effort is the development of a **National Dashboard**, an online supply chain traceability system designed to compile, synchronize, and verify data and maps for key commodities like palm oil, coffee, cocoa, and rubber at every stage of the supply chain.[14] This platform is envisioned as a comprehensive verification and data integration system.[17] Concurrently, efforts are underway to accelerate the issuance of Plantation Cultivation Registration Certificates (STDB) to independent farmers, aiming to address the land legality requirements of the EUDR.[16] Alongside governmental efforts, non-governmental organizations are also contributing. For instance, Kaoem Telapak has developed **Ground Truthed.id (GTID)**, a platform that facilitates bottom-up monitoring by enabling the collection of field-based evidence, geolocation data, and documentation of environmental violations through web and Android applications.[17] GTID is intended to complement official monitoring systems by providing independently verified ground-level information.[17]
The EUDR's stringent data demands, particularly for precise plot-level geolocation and end-to-end traceability [2], are proving to be a significant driver for change. Meeting these requirements with traditional, often paper-based or siloed data systems, is largely unfeasible for a nation with extensive agricultural lands and millions of small-scale producers.[16] This external regulatory pressure is

thus compelling Indonesia to accelerate investment in and development of national-level digital infrastructure, exemplified by the National Dashboard [14] and reforms in land registration systems like the STDB.[16] If these digital systems are developed with a focus on effectiveness, inclusivity, and sustainability, they hold the potential to yield broader benefits that extend beyond EUDR compliance. Such benefits could include strengthened national land governance, enhanced transparency across agricultural supply chains for all markets, more efficient resource management, and improved targeting of support for smallholder farmers. In this way, the EUDR, while posing immediate challenges, may act as an external catalyst for significant data modernization and systemic reforms within Indonesia's agricultural sector.

- The Imperative for Artificial Intelligence Solutions
  The vast scale of monitoring millions of hectares of diverse agricultural landscapes, tracking the provenance of commodities from a multitude of smallholder plots, and rigorously verifying deforestation-free claims necessitates the adoption of advanced technological solutions.9 Artificial Intelligence (AI), particularly its applications in analyzing remote sensing data (such as satellite imagery) and in sophisticated supply chain analytics, offers a powerful and potentially transformative toolkit to address these complex EUDR compliance challenges.9 AI-driven systems can automate the detection of deforestation events, assist in the accurate delineation of farm plot boundaries, enhance the sophistication of risk assessment models, and improve the overall efficiency and reliability of the due diligence processes mandated by the EUDR.
  While AI models can deliver powerful technical capabilities for monitoring and analysis [9], the context of EUDR compliance in Indonesia is deeply intertwined with the realities of its agricultural sector, which heavily involves smallholder farmers.[7] These smallholders often face limitations in terms of access to advanced technology, digital literacy, and may have legitimate concerns regarding data privacy, ownership, and the equitable distribution of benefits from new technologies.[5] Consequently, the successful deployment and sustained impact of AI-driven solutions for EUDR will depend not only on their technical sophistication and accuracy but critically on how these technologies are integrated into existing socio-economic structures, local governance, and agricultural workflows. A purely top-down, technology-centric approach to AI implementation, without due consideration for the social context, user needs (especially those of smallholders), capacity building, and trust, is likely to encounter significant adoption barriers or could even exacerbate existing inequalities.[22] This implies that the development of AI solutions should ideally be participatory, or at a minimum, involve rigorous validation with diverse stakeholder input. The solutions

should aim to be accessible, or their outputs should be translatable into actionable information for intermediaries and organizations supporting smallholders. Furthermore, ethical considerations, including data privacy, algorithmic bias, and benefit-sharing, must be proactively addressed. This frames the problem not merely as a technical challenge but as a socio-technical one, where the social and human dimensions are as critical as the algorithmic components. This thesis, while concentrating on the technical development of an AI model, will maintain an awareness of these broader factors, particularly in its discussion of implications and future work.

1.2. State of the Art (Theory)

The application of Artificial Intelligence (AI) to address environmental challenges and regulatory compliance is a rapidly evolving field. For the EUDR, AI offers potential solutions across deforestation monitoring, land use analysis, and supply chain verification.

AI, particularly machine learning (ML) and deep learning (DL) algorithms, has become increasingly central to environmental monitoring. These techniques are extensively used for analyzing vast quantities of satellite imagery, such as data from the Landsat and Sentinel missions, to perform tasks like Land Use/Land Cover (LULC) mapping, precise deforestation detection, and tracking various environmental changes over time.[30]

State-of-the-art deep learning architectures, including Convolutional Neural Networks (CNNs), U-Net and its sophisticated variants (e.g., Attention U-Net, Residual U-Net), and Transformer-based models, are at the forefront of image segmentation and object detection. These capabilities are directly applicable to the challenges of delineating agricultural plot boundaries and identifying deforestation patches from satellite imagery with high accuracy.[48] Hybrid models that combine the strengths of CNNs for local feature extraction with the global context modeling abilities of Transformers, such as the Siamese U-Net Transformer (SUT), are also showing significant promise in remote sensing change detection.[48]

Beyond geospatial analysis, AI techniques are being explored and deployed to enhance supply chain traceability and conduct sophisticated risk assessments. This includes the use of ML for detecting anomalies in complex supply chain data, which could indicate non-compliance or fraud, and AI for the automated verification of compliance documents.[10] Graph Neural Networks (GNNs) are an emerging area of interest for their ability to model the intricate interdependencies within supply chains and propagate risk signals effectively.[57]

Geospatial AI platforms, notably Google Earth Engine (GEE), provide powerful, cloud-based environments for accessing and analyzing extensive archives of Earth observation data at scale, facilitating the development and deployment of AI models for environmental applications.[65]

International organizations play a crucial role in this domain. The Food and Agriculture Organization (FAO) and the World Resources Institute (WRI) are actively involved in developing and promoting tools and platforms, such as FAO's SEPAL (System for Earth Observation Data Access, Processing and Analysis for Land Monitoring) and WRI's Global Forest Watch (GFW), which often incorporate AI and ML for forest monitoring and land assessment.[38] The United Nations also recognizes and advocates for the potential of AI in supporting climate action, including efforts to monitor and combat deforestation.[40]

Despite these advancements, significant challenges persist, particularly concerning the inclusion of smallholder farmers in these sophisticated technological systems. The specific needs and constraints of smallholders are widely acknowledged, prompting research and development into targeted AI solutions that are accessible, affordable, and effective in their context.[20]

The EUDR compliance landscape necessitates a convergence of AI applications in remote sensing with those in supply chain management. Effective compliance requires linking on-the-ground production activities, which can be monitored using remote sensing AI, with the intricate movements and documentation within the supply chain, areas where supply chain-focused AI can play a role.[23] While the literature shows distinct progress in both remote sensing AI and supply chain AI, a key frontier in the state of the art, and an area ripe for further contribution, is the effective and seamless integration of these two AI domains to create holistic, end-to-end solutions for EUDR. Although this thesis will concentrate on the remote sensing component, it is crucial to situate this work within the broader context of this necessary integration.

1.3. Gap Analysis

Despite the significant advancements in AI for environmental monitoring and the proliferation of tools and platforms, several critical gaps remain, particularly concerning the specific and granular requirements of the EUDR in the Indonesian context, especially for smallholder-dominated commodity sectors.

- **Generality of Existing Solutions and Lack of Plot-Level Specificity for EUDR:**

Many existing AI-driven deforestation monitoring systems, such as global platforms like Global Forest Watch [38], provide valuable large-scale assessments. However, they may lack the fine-grained resolution and specificity required for EUDR compliance at the individual plot level. The EUDR mandates traceability to the precise plot of land.[4] This is particularly challenging for commodities like coffee grown in complex agroforestry systems, which are common in Indonesia and are inherently more difficult to detect and delineate accurately using standard remote sensing approaches compared to monoculture plantations.[76] These systems often struggle with the spectral confusion between shaded coffee and natural forest canopy.

- **Challenge of Smallholder Plot Delineation in Complex Tropical Landscapes:** The accurate delineation of small (often less than 1-2 hectares), irregularly shaped, and often intercropped farm plots, characteristic of Indonesian smallholder agriculture, remains a formidable technical challenge, especially when relying on freely available medium-resolution satellite imagery like Sentinel-2 (10m resolution for key bands).[88] Existing datasets for training plot delineation models, such as AI4SmallFarms [89], while valuable, are primarily focused on mainland Southeast Asia (Vietnam/Cambodia) and may require significant adaptation, re-validation, or supplementation with Indonesia-specific data for coffee-growing regions.
- **Scarcity of Indonesia-Specific, Commodity-Focused Ground-Truth Data:** The development and validation of robust, localized AI models are heavily dependent on the availability of substantial, high-quality, and accurately labeled ground-truth data. While Indonesia possesses national mapping initiatives and geospatial data portals (e.g., INA-Geoportal, data from BRIN/LAPAN) [14], readily accessible, curated, and plot-level ground-truth datasets specifically designed for training AI models for EUDR-relevant coffee plot delineation and associated deforestation monitoring in Indonesia are not widely available or may be fragmented across different agencies and initiatives. The GTID initiative [25], while promising for bottom-up data collection, is relatively new, and the accessibility and format of its data for AI model training purposes are yet to be fully established.
- **Feasibility Constraints for Advanced AI Models within a Master's Thesis Scope:** Many cutting-edge AI models, such as very large Transformer architectures or comprehensive Graph Neural Networks for entire supply chains, demand extensive datasets, significant computational resources for training, and prolonged development times.[57] These requirements often make them less feasible for execution within the typical 6-month timeframe and resource constraints of a Master's thesis. There is a gap for proposing and demonstrating

an AI solution that is both state-of-the-art in its approach yet practically implementable and capable of delivering meaningful results within such constraints.

- **Operational Integration of AI Outputs into Due Diligence Workflows:** While AI can excel at detecting deforestation or mapping agricultural plots, the seamless and practical integration of these AI-generated outputs into the broader EUDR due diligence workflow—which includes risk assessment, verification of legal documents [28], and traceability reporting—remains an underexplored area in academic research that focuses on practical implementation pathways for Indonesian stakeholders, particularly for smallholders and their supporting organizations.
- **Specific AI Model Adaptation for Indonesian Coffee Agroforestry Systems:** Coffee in Indonesia is frequently cultivated in agroforestry systems, where coffee plants grow under the shade of various other trees. This creates a spectrally complex environment that makes it challenging to distinguish coffee plots from surrounding natural vegetation or other mixed cultivation using standard satellite image analysis techniques.[94] There is a specific research gap in developing and validating AI models that are explicitly adapted to delineate such coffee plots and monitor deforestation in their immediate vicinity with high accuracy.

The EUDR's core requirement of plot-level traceability and deforestation-free verification [4] highlights a critical "last mile" challenge. While global and national monitoring systems can provide valuable macro-level oversight and identify broad areas of concern, the regulation demands verification at the level of individual production units. Bridging the gap between large-scale AI-driven monitoring and the accurate identification, delineation, and deforestation status assessment of individual smallholder plots, especially those situated in complex, data-scarce environments like Indonesian coffee agroforestry systems, is paramount. It is precisely at this "last mile" that AI can offer significant practical contributions, but it is also where the most acute technical and data-related challenges lie. This thesis aims to address a key component of this "last mile" problem by developing a specialized AI model for plot delineation and associated deforestation risk assessment.

1.4. Problem Formulation (Research Question)

Given the challenges posed by the EUDR to Indonesian coffee producers, particularly smallholders, and the potential of AI in addressing these, this research seeks to answer the following primary questions:

How can a Siamese U-Net deep learning architecture, enhanced with attention mechanisms, be effectively designed and applied to multi-temporal Sentinel-2 satellite imagery and relevant ancillary geospatial data to:

1. Accurately detect deforestation events that have occurred after the EUDR cut-off date of December 31, 2020, in and around coffee-growing areas within selected regions of Indonesia?
2. Precisely delineate the boundaries of smallholder coffee plots situated within these complex agroforestry landscapes, distinguishing them from other land cover types?

Furthermore, can such a specifically designed AI model provide a technically feasible and practically valuable tool for Indonesian stakeholders, including small coffee farmers, cooperatives, and relevant government agencies, in supporting their due diligence processes and traceability requirements under the EUDR, with the development and initial validation being achievable within the timeframe of a 6-month Master's thesis?

1.5. Objective of Research (Research Objective, answer the Research Question)

The overarching goal of this research is to design, develop, and rigorously evaluate a novel Siamese U-Net deep learning model, incorporating attention mechanisms, for the dual and interconnected tasks of (a) detecting recent deforestation events and (b) delineating smallholder coffee plot boundaries in selected Indonesian coffee-growing regions, utilizing freely available Sentinel-2 satellite imagery. The ultimate aim is to demonstrate a proof-of-concept for an AI-driven tool that can support EUDR compliance efforts in Indonesia.

To achieve this primary objective, the following specific objectives are defined:

1. **Dataset Curation and Pre-processing:** To curate a comprehensive dataset for selected Indonesian coffee-growing regions. This will involve acquiring and pre-processing multi-temporal Sentinel-2 satellite imagery (spanning periods before and after the EUDR cut-off date of December 31, 2020). Ground-truth data for coffee plots and deforestation events will be sourced by adapting existing relevant datasets (e.g., AI4SmallFarms [89], with validation for Indonesian coffee context) and/or through limited, targeted manual annotation of Very High-Resolution (VHR) satellite imagery for validation purposes.
2. **AI Model Design and Implementation:** To design and implement a Siamese U-Net deep learning architecture. This architecture will feature two identical encoder pathways for processing paired temporal images (for change detection) and will incorporate attention mechanisms (e.g., spatial and/or channel attention)

within its decoder structure. The design will be tailored to effectively identify subtle changes indicative of deforestation and to segment small, often irregularly shaped coffee plots from spectrally complex Sentinel-2 data.

3. **Model Training and Validation:** To train the proposed Siamese U-Net model with attention mechanisms using the curated dataset. Its performance will be rigorously validated and benchmarked against a baseline U-Net model (or a standard change detection method) trained on the same dataset, employing appropriate and clearly defined accuracy metrics (e.g., F1-score, Intersection over Union (IoU), precision, recall).
4. **Utility Assessment for EUDR Compliance:** To analyze the outputs of the trained model (i.e., deforestation maps and delineated plot boundaries) to assess their potential utility in providing credible evidence for deforestation-free claims and supporting plot-level traceability, as mandated by the EUDR.
5. **Feasibility and Contextual Discussion:** To discuss the practical feasibility of deploying such an AI model within the Indonesian context, considering the operational constraints and the 6-month timeframe of the Master's thesis. The discussion will also cover the model's potential contribution to informing national EUDR mitigation strategies and supporting sustainable coffee production.

1.6. Limitation

This research, while ambitious, is subject to several limitations inherent in a Master's thesis project:

- **Geographic Scope:** The empirical investigation will be confined to a limited number of selected coffee-growing regions or districts in Indonesia (e.g., one or two representative areas such as Gayo in Aceh, or a district in Lampung or Toraja, contingent on data accessibility and the prevalence of smallholder coffee farming). Consequently, the direct generalizability of the findings to all diverse coffee-producing landscapes across Indonesia will be an extrapolation, and the model might require further adaptation for different agro-ecological zones.
- **Commodity Focus:** The research will concentrate exclusively on coffee as the target commodity. While the developed AI methodology for plot delineation and deforestation detection might possess transferable elements, its direct applicability and performance for other EUDR-affected commodities prevalent in Indonesia (such as palm oil, rubber, or cocoa, which have distinct spectral signatures and cultivation patterns) will not be empirically tested within this study.
- **Data Availability and Ground Truth:** The accuracy, comprehensiveness, and accessibility of ground-truth data for training and, more critically, for robust validation will be a significant constraint. The thesis will prioritize the use of

publicly available Sentinel-2 satellite imagery. For ground truth, it will aim to leverage and adapt existing datasets (e.g., AI4SmallFarms [89], if applicable to coffee and adaptable to Indonesian context) or conduct very limited, targeted manual annotation of VHR imagery primarily for creating a high-quality, independent test/validation set. The creation of a large-scale, new ground-truth dataset is beyond the scope and timeframe of a 6-month thesis.[88] This may impact the scale of training and the statistical power of the validation.

- **Model Complexity and Computational Resources:** The complexity of the AI model architecture and the extent of hyperparameter tuning will be constrained by the available computational resources (e.g., university high-performance computing clusters, services like Google Colab Pro) and the 6-month project timeline allocated for model development, training, and evaluation.
- **Scope of EUDR Compliance:** This research focuses on the technical geospatial aspects of EUDR compliance, specifically deforestation detection and plot delineation. It will not encompass other critical components of the due diligence process, such as the full verification of legality (which involves analyzing land tenure documents, permits, etc.) or the development of a complete end-to-end supply chain traceability system. However, the outputs are intended to provide crucial data inputs for these broader compliance activities.
- **Temporal Scope of Deforestation Detection:** While the objective is to detect deforestation occurring after December 31, 2020, the practical ability to do so will depend on the availability of consistent, cloud-free Sentinel-2 imagery for both the pre- and post-cutoff periods in the selected study areas. Gaps in historical imagery or ground truth may limit the temporal depth and comprehensiveness of the change analysis.

1.7. Hypothesis (Optional)

The following hypotheses will guide the experimental phase of this research:

- **H1:** A Siamese U-Net architecture incorporating attention mechanisms will achieve a statistically significant improvement in key performance metrics (specifically, F1-score and Intersection over Union - IoU) for both (a) the detection of deforestation events and (b) the delineation of smallholder coffee plot boundaries within Indonesian coffee agroforestry landscapes, when applied to Sentinel-2 satellite imagery, as compared to a standard U-Net architecture trained for the same tasks.
- **H2:** The AI model developed in this research can identify deforestation events linked to coffee production areas and delineate smallholder coffee plots with a level of accuracy and reliability that provides valuable, actionable information to

support EUDR due diligence processes for smallholders and their intermediaries in the selected Indonesian study areas.

2. Detail Theory / Literature Review

This chapter reviews the existing body of knowledge pertinent to the EUDR, AI applications in deforestation and LULC monitoring, AI for supply chain management, relevant initiatives by governmental and non-governmental organizations, and specific considerations for AI in smallholder agriculture.

2.1. The EU Deforestation Regulation (EUDR): Mandates and Implications for Indonesia

The EUDR (Regulation (EU) 2023/1115) establishes stringent requirements for operators placing specific commodities on the EU market or exporting them. These commodities—cattle, cocoa, coffee, oil palm, rubber, soya, and wood, along with numerous derived products listed in Annex I of the regulation (e.g., meat products, leather, chocolate, coffee, palm nuts, palm oil derivatives, natural rubber products, soybeans, wood products) [1]—must be proven 'deforestation-free', legally produced, and covered by a due diligence statement.[1] 'Deforestation-free' means that the land on which the commodity was produced was not converted from forest to agricultural use after December 31, 2020. For wood products, it also means harvesting did not induce forest degradation after this date.[2] 'Forest degradation' is defined as structural changes to forest cover, taking the form of the conversion of primary forests or naturally regenerating forests into plantation forests or other wooded land.[2]

The due diligence process is central to the EUDR and consists of three main steps [2]:

1. **Information Gathering (Article 9):** Operators must collect comprehensive and verifiable information. This includes the commodity description, quantity, country of production, and crucially, the geolocation coordinates (latitude and longitude) of all plots of land where the commodity was produced.[2] For products containing or made using cattle, this includes geolocation of all establishments where cattle were kept. Evidence must also be gathered to demonstrate that the commodity is deforestation-free and that it was produced in accordance with the relevant legislation of the country of production.
2. **Risk Assessment (Article 10):** Operators must analyze the collected information to assess the risk of non-compliance. This assessment must consider criteria such as the country's risk classification (low, standard, or high, to be determined by the EU Commission; until then, all are standard risk [2]), the presence of forests and deforestation in the production area, the type of commodity, information

from suppliers (including smallholders), the complexity of the supply chain, and the risk of mixing with products of unknown origin or non-compliant products.[2]

3. **Risk Mitigation (Article 11):** If the risk assessment indicates a non-negligible risk, operators must adopt and implement procedures and measures to reduce that risk to a negligible level. These measures can include requesting additional information, data, or documents, carrying out independent surveys or audits, or providing support (e.g., capacity building, investments) to suppliers, particularly smallholders, to achieve compliance.[2]

The EUDR also mandates compliance with the **legality requirement**, meaning commodities must be produced in accordance with all relevant applicable laws in the country of production. This covers laws related to land use rights, environmental protection, forest-related rules (including forest management and biodiversity conservation where directly related to harvesting), third parties' rights (including rights of Indigenous Peoples and local communities, and the principle of Free, Prior and Informed Consent - FPIC), labor rights, human rights protected under international law, tax legislation, anti-corruption laws, and trade and customs regulations.[2]

The implications for Indonesia, a major global supplier of several EUDR-affected commodities, are substantial. The palm oil sector, where Indonesia is the world's largest producer and a key exporter to the EU [16], faces immense pressure. Millions of smallholders are involved in palm oil production, and ensuring their compliance with EUDR's traceability and legality requirements is a significant challenge.[5] Similar challenges exist for other commodities like coffee and cocoa, which are also predominantly grown by smallholders, often in complex agroforestry systems that make plot delineation and deforestation monitoring inherently difficult.[17] The timber and rubber sectors also face comparable hurdles in verifying traceability and legality.[17]

Smallholder farmers in Indonesia encounter specific difficulties in meeting EUDR standards. These include limited access to technology for GPS mapping and digital record-keeping, insufficient financial resources to cover costs associated with certification or traceability systems, lack of formal land titles or clear land tenure documentation, and navigating complex supply chains with multiple intermediaries.[5] The proposed AI solution in this thesis must be developed with an awareness of these socio-economic realities, aiming to provide outputs that can eventually be made accessible or usable by organizations supporting these smallholders.

The EUDR's stringent demand for precise geolocation of all production plots [3] emerges as the most data-intensive and technically challenging aspect for widespread compliance, particularly for smallholders who may manage multiple small, dispersed plots. This makes the tasks of accurate plot delineation and linking these plots to deforestation events (or the absence thereof) the most critical areas where AI can offer a scalable and effective solution. This thesis directly targets this central requirement by focusing on AI-driven plot delineation and deforestation risk assessment.

It is also important to note the EUDR's definition of 'forest degradation' for wood products, which involves the conversion of primary or naturally regenerating forests into plantation forests or other wooded land, or harvesting practices that induce degradation.[2] This is a more nuanced concept than clear-cut deforestation and may necessitate different AI monitoring approaches, potentially involving the detection of subtle changes in forest structure or canopy density. While this thesis focuses on deforestation related to coffee, this distinction is relevant for the broader scope of AI research aimed at comprehensive EUDR compliance.

2.2. AI for Deforestation Monitoring and Land Use/Land Cover (LULC) Change Detection

The use of AI, particularly in conjunction with remote sensing data, has revolutionized the way deforestation and LULC changes are monitored globally and locally.

- Remote Sensing Data Sources:
  A variety of satellite data sources are available for LULC mapping and change detection, each with its own characteristics:
  - **Optical Satellites:**
    - **Sentinel-2 (ESA):** This mission provides multispectral imagery (MSI) with 13 spectral bands at spatial resolutions of 10m, 20m, and 60m, and a high revisit frequency of approximately 5 days globally with two satellites.[32] Its free data policy, global coverage, and suitable resolution make it a primary choice for LULC classification, vegetation monitoring, and deforestation detection, and it is the main data source for this thesis.
    - **Landsat (NASA/USGS):** With an archive spanning several decades, Landsat offers imagery at a 30m spatial resolution, making it invaluable for long-term historical deforestation analysis and establishing baseline forest cover.[38] This can be crucial for verifying the EUDR's 2020 cut-off date.
  - **SAR (Synthetic Aperture Radar) Satellites:** Missions like Sentinel-1 (ESA) and ALOS PALSAR-2 (JAXA) utilize radar technology, which has the significant advantage of being able to penetrate cloud cover. This capability is particularly crucial for consistent monitoring in persistently cloudy tropical

regions like Indonesia.[31] While SAR data is not the primary focus of this thesis, its potential to complement optical data in areas with high cloud obscuration is acknowledged.
    - **Very High-Resolution (VHR) Commercial Imagery:** Data from commercial providers like Planet and Maxar offer spatial resolutions of less than 5 meters, enabling the detection of very fine details. This is useful for validating medium-resolution analyses and for mapping very small or complex features.[31] However, the cost and limited accessibility of VHR data for large areas make it more suitable for targeted validation within the scope of a Master's thesis rather than for extensive training or primary analysis.
- Deep Learning Models for Geospatial Image Analysis:
  Deep learning has emerged as the dominant approach for analyzing satellite imagery for LULC classification and change detection.
    - **Convolutional Neural Networks (CNNs):** These are foundational architectures for image-based tasks, excelling at automatically learning hierarchical spatial features from raw pixel data. Various CNN architectures, such as ResNet, have been successfully applied to remote sensing image analysis.[42]
    - **U-Net Architecture and Variants:** The U-Net, characterized by its encoder-decoder structure with skip connections, has proven highly effective for semantic segmentation (pixel-wise classification) in both biomedical and remote sensing imagery. The skip connections allow the decoder to utilize fine-grained feature maps from the encoder, which helps in preserving spatial details and achieving precise boundary delineation—a critical capability for plot mapping.[48]
        - **Attention U-Net:** This variant incorporates attention mechanisms into the U-Net architecture. Attention mechanisms enable the model to selectively focus on the most relevant features or regions within an image, which can improve performance, especially when dealing with small, complex, or partially obscured objects like smallholder plots in agroforestry systems or subtle signs of early-stage deforestation.[48] This is a key component of the model proposed in this thesis.
        - **Siamese Networks:** These networks consist of two (or more) identical subnetworks (often CNNs) that process different input samples (e.g., image patches from two different time points). The outputs of these subnetworks are then compared to learn a similarity or difference metric. Siamese architectures are inherently well-suited for change detection tasks, as they can directly learn to identify differences between temporal image pairs.[48] The proposed model will leverage a Siamese U-Net

structure.

  ○ **Transformer Models:** Originally developed for Natural Language Processing (NLP), Transformer models, particularly Vision Transformers (ViTs) and their derivatives (e.g., Swin Transformer), are increasingly being adapted for computer vision tasks. Their self-attention mechanism allows them to effectively model long-range dependencies and capture global context within an image, which can be advantageous for analyzing large and complex scenes.[42]
  ○ **Hybrid CNN-Transformer Models:** Recent research explores combining the local feature extraction capabilities of CNNs with the global context modeling strengths of Transformers. Models like the Siamese U-Net Transformer (SUT) have demonstrated promising results for remote sensing change detection, suggesting that such hybrid approaches can offer superior performance.[48] While implementing a full SUT model might be too complex for a 6-month thesis, the proposed Siamese U-Net with attention mechanisms represents a step towards leveraging similar principles of combining robust feature extraction with focused attention.

- Comparative Analysis of Models for Deforestation/Plot Delineation:
  When selecting an AI model for tasks like deforestation detection and plot delineation, several factors must be considered, including model complexity, data requirements, computational cost, and expected performance.

Feature	Random Forest (RF) / SVM	U-Net (and variants)	Transformer-based Models
Primary Strength	Robust on tabular/moderate-dimension data, less tuning for RF	Excellent for image segmentation, good boundary localization	Strong global context modeling, good for complex relationships
Data Requirements	Moderate	Moderate to Large (can work with smaller sets than Transformers)	Large to Very Large
Computational Cost	Low to Moderate	Moderate to High (especially with deeper variants)	High to Very High
Hyperparameter	RF: Minimal; SVM:	Moderate	Significant

Tuning	Significant		
Feature Engineering	Often requires manual feature engineering	Learns features automatically (less engineering needed)	Learns features automatically (less engineering needed)
Interpretability	RF: Moderate; SVM: Low (non-linear)	Low to Moderate (attention can add some interpretability)	Low
Performance (Images)	Baseline, can be good with engineered features	State-of-the-art for many segmentation tasks	State-of-the-art, especially with large datasets
Suitability for Thesis	Good as baseline	High (feasible, good performance)	Potentially too complex/data-hungry for 6 months

Table 2.1: Comparative Analysis of AI Model Architectures for Geospatial Tasks. (Sources: [43, 48, 50, 51, 52, 53, 54, 55, 101, 102, 106, 107])

Traditional machine learning algorithms like Random Forest (RF) and Support Vector Machines (SVM) have been widely used and can serve as important baselines.[43, 53, 101, 102, 106, 107] RF is known for its robustness and ability to handle high-dimensional data with relatively little hyperparameter tuning, while SVMs can be powerful but often require more careful parameter selection. However, for complex image segmentation tasks, deep learning models like U-Net generally outperform these traditional methods by automatically learning intricate spatial patterns.[53]

U-Net architectures are particularly well-suited for semantic segmentation of satellite imagery and can achieve good results even with moderate-sized datasets, a key consideration for a time-constrained thesis.[43, 52, 53, 55, 91] Transformer-based models, while powerful in capturing global context and achieving state-of-the-art results on large benchmark datasets, are typically more data-hungry and computationally intensive to train, which might pose feasibility challenges for a 6-month Master's project.[48, 50, 51, 101, 102] The proposed approach of using a Siamese U-Net with attention mechanisms aims to strike a balance, leveraging the proven strengths of U-Net for segmentation and Siamese networks for change detection, while enhancing performance through attention, all within a practically achievable framework.

The challenge of accurately delineating small, irregular plots, especially in heterogeneous tropical smallholder landscapes, is well-documented.[88, 89] Attention mechanisms, when integrated into architectures like U-Net, offer a promising avenue to address this. By enabling the model to assign greater importance to salient features or regions within the input image or intermediate feature maps, attention can help in better distinguishing these challenging features from their complex surroundings.[50, 51, 52, 53] This targeted focus is hypothesized to improve the delineation accuracy for smallholder coffee plots and the detection of subtle deforestation signals.

Furthermore, the EUDR's strict deforestation cut-off date of December 31, 2020 [2] makes change detection a core requirement. Siamese network architectures are inherently designed for comparing two inputs, such as satellite images captured before and after this cut-off date.[48, 51] This architectural choice directly aligns with the fundamental logic of identifying changes that have occurred *after* a specific point in time, offering a more direct and potentially more effective approach for this specific EUDR task compared to methods that rely on single-image classification or post-classification comparison.

2.3. AI for Supply Chain Traceability and Risk Assessment (Brief Overview - Secondary Focus for this Thesis)

While this thesis primarily focuses on the geospatial AI aspects of deforestation detection and plot delineation, it is important to acknowledge the broader role of AI in ensuring end-to-end EUDR compliance, which involves supply chain traceability and comprehensive risk assessment.

Technologies like **blockchain** are frequently cited for their potential to create transparent and immutable records of transactions and product movements within supply chains.[15] AI can complement blockchain systems by analyzing the data for anomalies, verifying the authenticity of information being recorded, or predicting potential risks based on patterns in the transactional data.

AI-driven **risk assessment models**, often employing machine learning techniques, can analyze a wide array of data sources. These include supplier information, historical transaction patterns, regional deforestation rates, socio-economic indicators, and even news or social media sentiment related to specific suppliers or regions, to predict the likelihood of compliance risks.[10]

Graph Neural Networks (GNNs) are emerging as a particularly powerful tool for modeling the complex, interconnected nature of modern supply chains.[57] GNNs can represent supply chain entities (e.g., farms, processors, exporters, plots of land) as nodes and their relationships (e.g., transactions, transportation links, geographical proximity) as edges. This allows them to learn how risks (such as deforestation or illegal sourcing) might propagate through the network and to identify critical vulnerabilities. Research by Wasi et al. (2024) indicates that GNN-based models can outperform traditional ML and other DL models in supply chain anomaly detection tasks by a significant margin (15-40%).[62] However, challenges related to the scalability of GNNs for very large global supply chains and the interpretability of their decisions remain active areas of research.[57]

AI, specifically **Natural Language Processing (NLP)** techniques, also holds potential for automating the review and verification of compliance-related documents, such as land tenure certificates, harvesting permits, or labor contracts.[28] For instance, CoolX, a commercial EUDR solution provider, mentions using AI for reviewing documents related to land usage rights.[28]

Effective EUDR risk assessment ultimately depends on the ability to fuse diverse data types. It cannot rely solely on satellite-derived geospatial data. A holistic approach requires the integration of outputs from geospatial AI models (like the one proposed in this thesis, which can provide plot delineations and deforestation alerts) with supply chain transactional data, supplier profiles, and information extracted from legal and compliance documents. The geospatial outputs of this research are therefore envisioned as critical inputs for more comprehensive AI-driven risk assessment systems, potentially leveraging GNNs or other advanced analytical platforms.[10]

2.4. AI Applications for EUDR/Deforestation by Governments and Organizations

Various governmental bodies, international organizations, and commercial entities are actively developing and deploying AI-driven solutions to address deforestation and support compliance with regulations like the EUDR.

- **Indonesian Initiatives:**
    - **National Dashboard:** The Indonesian government is spearheading the development of a National Dashboard, envisioned as a centralized supply chain traceability system. This platform aims to compile, synchronize, and verify data and maps for key commodities like palm oil, coffee, cocoa, and

rubber, functioning as a national verification and integration system to support EUDR compliance.[14] While technical details are still emerging, its role in consolidating diverse data sources is critical.
    - **Ground Truthed.id (GTID):** Developed by the NGO Kaoem Telapak, GTID is a bottom-up monitoring platform that utilizes field-based evidence collection, including geolocation data and photographic documentation of environmental violations. It operates through web and Android applications, supports offline data capture, and includes an internal verification system. GTID is designed to complement official satellite monitoring efforts by providing ground-level validation and context.[17] The platform aims for API interoperability, which could allow its data to be integrated with other systems.[25]
    - **INA-Geoportal & LAPAN/BRIN:** Indonesia's national geospatial data infrastructure, including the INA-Geoportal and data services from the National Research and Innovation Agency (BRIN, which absorbed LAPAN, the former space agency), provides foundational satellite data (e.g., Landsat, SPOT, Pleiades, TerraSAR-X) and land cover information. BRIN offers pre-processed data, including geometric and radiometric corrections, which are essential inputs for any AI-based monitoring system operating in Indonesia.[97]
- **International Examples:**
    - **Brazil (INPE - PRODES & DETER):** Brazil has a long history of satellite-based deforestation monitoring through its National Institute for Space Research (INPE). The PRODES system provides annual deforestation assessments for the Amazon, while the DETER system offers near real-time alerts of deforestation activities.[85] INPE is currently enhancing these systems by incorporating AI techniques (both supervised and unsupervised segmentation using Sentinel-2 imagery, Spectral Mixture Models, Random Forest, and Self-Organizing Maps) to supplement the Prodes historical series for creating an EUDR-compliant landmark deforestation map for Brazil.[104] Key challenges include persistent cloud cover, for which Sentinel-1 SAR data is used as a solution, and the alignment of historical Landsat data (30m) with newer Sentinel-2 data (10m).[104] The DETER-R system, which uses Sentinel-1 SAR, has demonstrated high accuracy in providing near real-time deforestation alerts with very low false positive rates.[103]
    - **FAO (Food and Agriculture Organization of the United Nations):** The FAO promotes various tools and initiatives for forest monitoring. A key platform is SEPAL (System for Earth Observation Data Access, Processing and Analysis for Land Monitoring), an open-source, cloud-based platform that facilitates access to and processing of satellite data for forest and land monitoring,

potentially incorporating AI/ML capabilities.[72] The FAO also works on developing data sharing protocols and enhancing traceability in agricultural supply chains to support regulations like the EUDR.[76]
    - **WRI (World Resources Institute) - Global Forest Watch (GFW):** GFW is a widely utilized online platform that provides open access to data, analytical tools, and alerts on global forest change. It primarily uses Landsat and other satellite data, processed with supervised learning algorithms, to track tree cover loss.[38] GFW Pro is a service specifically aimed at companies for assessing deforestation risk in their supply chains.[73] While GFW provides valuable global insights, its standard products (e.g., 30m resolution tree cover loss) may have limitations for the granular plot-level verification required by EUDR, and its definition of "tree cover loss" does not always equate to "deforestation" under EUDR terms.[38]
    - **UN Initiatives:** The United Nations Framework Convention on Climate Change (UNFCCC) and the UN Environment Programme (UNEP) are exploring and promoting the use of AI for climate action. This includes applications in predicting deforestation rates using neural networks trained on spatial-temporal data, with capabilities for continuous model retraining as new data becomes available.[40]
- **Commercial AI Solutions for EUDR:** A growing number of private companies are offering AI-powered SaaS (Software as a Service) platforms designed to help businesses comply with the EUDR. These solutions typically combine satellite imagery analysis (often Sentinel-2), supply chain mapping tools, automated risk assessment algorithms, and due diligence reporting functionalities. Examples include CoolX [28], Picterra [9], TraceX [23], IntegrityNext [10], LiveEO TradeAware [11], CropIn [30], and DISS-CO.[13] These platforms often highlight AI capabilities for land cover analysis, automated document review [28], geospatial AI for plot verification and deforestation detection using Sentinel-2 imagery [9], and integration with enterprise resource planning (ERP) systems.

The landscape of AI for EUDR compliance reveals a dichotomy between large-scale, often satellite-driven, "top-down" monitoring systems (e.g., GFW, PRODES, the concept of Indonesia's National Dashboard) [17] and more localized, community-focused, field-data-driven "bottom-up" systems (e.g., GTID).[25] Achieving robust EUDR compliance, especially in complex smallholder landscapes, will likely necessitate a synergistic combination of these approaches. Top-down AI systems can efficiently identify potential risk areas and monitor large-scale changes, providing an initial layer of scrutiny. Bottom-up data collection, potentially enhanced by AI-assisted field tools, can then provide crucial validation, contextual information, and detailed verification, particularly for individual smallholder plots and in areas with complex land tenure or agroforestry practices. This thesis focuses on developing a top-down AI methodology (satellite image analysis), but its outputs are envisioned as being highly valuable for subsequent verification or integration with bottom-up data streams or systems.

- **Table 2.2: Overview of Selected AI-driven EUDR Compliance/Deforestation Monitoring Initiatives**

Organization/Country/Company	AI Techniques Used/Platform	Key Focus	Commodity Focus (if any)	Reported Outcomes/Lessons Learned	Relevance to Indonesian EUDR
Indonesia (Govt.)	National Dashboard (in development) - Data integration, verification	Supply chain traceability, data synchronization	Palm oil, coffee, cocoa, rubber, etc.	Aim: comprehensive national system.[14]	Directly aims to support Indonesian producers and exporters in meeting EUDR requirements.
Kaoem Telapak (NGO)	Ground Truthed.id (GTID) - Mobile/Web app, geolocation, field evidence, internal verification	Bottom-up monitoring, documenting violations, complementing satellite systems	All EUDR commodities	Focus on verifiable ground-truth, offline capability, API for interoperation.[25]	Provides a potential source of validation data and a model for community-based monitoring relevant to smallholders.
Brazil (INPE)	PRODES/DETER - Satellite imagery (Landsat, Sentinel-2, Sentinel-1	Deforestation alerts (DETER), annual deforestation	General deforestation (Amazon, other biomes)	High accuracy for alerts (DETER-R <0.2% false positives	Demonstrates successful large-scale AI-based deforestation monitoring

	SAR), AI (segmentation, RF, SOM) for EUDR Landmark Map	assessment (PRODES)		[103]). Challenges with cloud cover (SAR as solution), aligning historical data.[104]	in a tropical country with similar challenges (clouds, small-scale deforestation). Use of SAR is a key lesson.
FAO	SEPAL Platform - Cloud computing, satellite data processing (potential AI/ML modules)	Forest & land monitoring, capacity building	General forest/land	Open-source, enables countries to use EO data for climate action.[72]	Provides tools and frameworks that could be leveraged by Indonesian institutions for national monitoring.
WRI	Global Forest Watch (GFW) / GFW Pro - Satellite imagery (Landsat), supervised learning, near real-time alerts	Global forest change monitoring, supply chain risk assessment (GFW Pro)	General forest, specific commodity tools	Widely used, provides global data. Limitations: 30m resolution, tree cover loss vs. deforestation definition.[38]	Offers valuable global context and alert systems, but may need supplementation with higher-resolution, locally calibrated models for plot-level EUDR compliance in Indonesia.
CoolX (Commercial)	AI for remote sensing land cover analysis, AI for	EUDR compliance for importers/exporters,	Coffee, cocoa, soy, palm oil, etc.	Case study with Caravela Coffee: analyzed 150	Demonstrates commercial application of AI for both geospatial

	document review (land rights, labor contracts)	deforestation-free supply chains		smallholder farms, validated documents.[28]	analysis and document review relevant to EUDR.
Picterra (Commercial)	Geospatial AI, supplier data verification, automated risk assessment, deforestation monitoring using approved/custom maps	EUDR compliance, supply chain mapping, automated reporting to TRACES	Coffee, cocoa, wood, palm oil, soy, etc.	Digitalizes sourcing parcels, >30 stress tests for data quality, automated DDS reports.[9]	Offers insights into commercial approaches for data quality assurance and automated reporting, relevant for scalable solutions.
IntegrityNext (Commercial)	AI-powered risk analysis (Sentinel-2, land classification, change detection), geolocation verification, API to TRACES	Automated due diligence, supplier data verification, EUDR compliance	Soy, palm oil, cocoa, rubber, coffee, etc.	Automated DDS submission, ERP integration.[10]	Highlights the integration of AI with regulatory reporting systems (TRACES) and enterprise systems.
LiveEO (TradeAware)	AI-powered satellite analytics (high-res), legality checks by local experts, ERP integration	End-to-end EUDR compliance, deforestation assessment, legality assessment	All EUDR commodities	Claims up to 98% error reduction vs. open-source data, automated checks in 80+ countries.[11]	Emphasizes high-accuracy proprietary analytics and integration of legal expertise, potentially offering higher precision than generic

					models.

This table provides a structured overview of existing AI applications for EUDR and deforestation monitoring, demonstrating awareness of current efforts and helping to identify common approaches, successful strategies (e.g., Brazil's use of SAR for cloud cover [103, 104]), and specific gaps this thesis can address by comparing its proposed approach.

2.5. AI for Smallholder Agriculture and Inclusion

The EUDR's impact on smallholder farmers is a significant concern, and AI presents both opportunities and challenges in this context.

- **Specific Challenges for Smallholders with EUDR:** Beyond the general compliance burdens of traceability and legality, smallholders face unique difficulties. These include the accurate mapping of their plots, which are often small (frequently less than one or two hectares), irregularly shaped, and may be part of complex intercropped or agroforestry systems.[88] They often lack digital records of their farm boundaries or production, have limited access to or familiarity with advanced AI tools or the data they require, and may struggle with the costs associated with certification or technology adoption.[7]
- **AI for Smallholder Plot Delineation:**
    - **Datasets:** The development of AI models for smallholder plot delineation relies heavily on suitable training datasets. The **AI4SmallFarms** dataset, covering parts of Vietnam and Cambodia, comprises 439,001 field polygons digitized from satellite images and provides Sentinel-2 imagery, with U-Net models benchmarked on it.[89] Another large dataset, **FBIS-22M**, focuses on European agricultural fields and includes 22 million instance masks from various satellite sources, including VHR imagery.[88] These examples underscore the need for large, diverse, and accurately labeled datasets. For this thesis, while AI4SmallFarms is geographically relevant to Southeast Asia, its direct applicability to Indonesian coffee agroforestry systems would need careful validation, and it may serve as a methodological reference or a source for transfer learning if feasible.
    - **Methods:** Common AI methods for plot delineation include Fully Convolutional Networks (FCNs), U-Net architectures, and post-processing techniques like contour closing and polygonization.[88] Key challenges identified

in the literature include the small average size of crop fields in smallholder systems and discrepancies between the resolution of available satellite imagery (like Sentinel-2) and the scale of these features, which can lead to difficulties in precise boundary delineation.[88]

- **AI for Smallholder Engagement and Support:** To make AI-driven solutions beneficial for smallholders, platforms are emerging that combine geospatial tools with farmer engagement features. For example, **Mergdata by Farmerline** utilizes mobile GPS mapping for farmer registration and plot delineation (Collect MD), satellite monitoring for land history verification (Terra MD), and an AI-powered multilingual chatbot (Darli AI) to provide advice, alerts, and support to farmers in local languages.[20] This demonstrates a pathway for translating complex AI outputs into usable information and support mechanisms for smallholders.
- **Ground Truth Collection in Smallholder Systems:** Acquiring accurate ground-truth data in smallholder landscapes is labor-intensive and often forms a bottleneck for training supervised AI models. Methodologies typically involve ground-based surveys using GPS devices to record plot boundaries and crop types, participatory mapping approaches involving local communities, and manual digitization of plot boundaries from VHR satellite imagery.[90] Given the constraints of a Master's thesis, large-scale primary ground-truth collection is not feasible; hence, the strategy will focus on leveraging existing datasets or conducting very targeted validation exercises.
- **Ethical Considerations:** The application of AI in smallholder agriculture, particularly for regulatory compliance like EUDR, raises important ethical questions. These include data privacy (especially concerning farm locations and personal information), data ownership and governance, the potential for algorithmic bias to disproportionately affect vulnerable groups, the risk of exclusion if AI systems are not designed inclusively or if smallholders lack the capacity to engage with them, and ensuring that the benefits of these technologies are shared equitably and contribute to improved livelihoods.[5] The EUDR itself acknowledges the need for partnerships and support for smallholders.[8] Any AI solution developed should be mindful of these ethical dimensions, aiming for transparency and fairness.

For EUDR compliance, there is a symbiotic relationship between plot delineation and deforestation monitoring, especially for smallholders. It is insufficient to merely detect a deforestation event; the EUDR requires linking this event (or its absence) to a specific plot of land where a commodity was produced.[6] In smallholder systems, where plots are numerous, often contiguous, and intermingled with other land uses, accurate AI-driven plot delineation becomes a critical prerequisite for conducting a reliable deforestation risk assessment associated with their specific production units. An AI model that can effectively perform both tasks, or where the output of plot delineation directly informs the deforestation monitoring process, is therefore highly valuable. This thesis proposes a dual-task approach to address this interconnected challenge.

2.6. Identification of Gaps and Justification for Proposed Work

Based on the preceding review of literature and existing solutions, several key gaps can be identified, justifying the proposed research:

- **Recap of Identified Gaps:**
    1. **Lack of Precision for Smallholder Coffee Plots in Indonesian Agroforestry:** Existing large-scale deforestation monitoring systems and generic LULC classification models often lack the precision and specificity required for accurately delineating smallholder coffee plots within the complex and heterogeneous agroforestry systems prevalent in Indonesia, especially using freely available medium-resolution Sentinel-2 data.
    2. **Need for Integrated Plot-Deforestation Linkage:** While some tools focus on deforestation and others on plot mapping, there is a need for AI models that can effectively link deforestation events (or their absence post-December 31, 2020) directly to these specifically delineated smallholder coffee plots, which is crucial for EUDR's plot-level due diligence.
    3. **Limited Indonesia-Specific, Coffee-Focused, Plot-Level Ground-Truth Data:** The availability of comprehensive, curated, and publicly accessible ground-truth datasets for training and rigorously validating AI models for coffee plot delineation and associated deforestation in the specific context of Indonesian agroforestry landscapes is limited. This hinders the development of locally optimized and robust models.
    4. **Feasibility of SOTA Solutions for Master's Thesis:** While highly sophisticated AI architectures (e.g., very large Transformers, end-to-end GNNs for entire supply chains) are being researched, there is a gap for proposing and demonstrating a solution that leverages state-of-the-art principles (like attention and Siamese networks) but remains feasible to implement, train, and evaluate within the typical resource and time constraints of a Master's thesis, while still offering a significant technical contribution.
    5. **Under-Explored Application of Siamese U-Net with Attention for EUDR in Indonesia:** The specific application of a Siamese U-Net architecture, enhanced with attention mechanisms, to the dual problem of deforestation detection and smallholder coffee plot delineation using Sentinel-2 data within

the Indonesian EUDR compliance context represents an underexplored area of research.

- Justification for the Proposed Work:
  The proposed research is justified by its potential to address these identified gaps through a novel, significant, timely, and feasible approach:
  - **Novelty:** The core novelty lies in the specific design and application of a Siamese U-Net architecture incorporating attention mechanisms, tailored for the dual tasks of deforestation detection and smallholder coffee plot delineation, specifically within the challenging context of Indonesian coffee agroforestry systems and the regulatory demands of the EUDR, using Sentinel-2 data. While the constituent components (Siamese networks, U-Nets, attention mechanisms) are known in the broader AI and remote sensing fields, their synergistic combination and application to this precise problem domain and geographic focus is novel and warrants investigation.
  - **Significance:** An accurate, reliable, and feasible AI tool for these tasks would be highly significant for Indonesian stakeholders. This includes government agencies (e.g., Ministry of Agriculture, Ministry of Environment and Forestry, the national EUDR Task Force), farmer cooperatives, exporters, and ultimately, smallholder coffee farmers themselves. Such a tool could substantially aid in meeting the EUDR's stringent due diligence requirements, potentially reducing compliance costs, minimizing the risk of market exclusion for smallholders, and contributing to more sustainable coffee production practices and forest conservation efforts in Indonesia.
  - **Timeliness:** With the EUDR implementation deadlines approaching (December 2025 for larger operators [1]), there is an urgent and pressing need for practical, implementable, and effective solutions that can support compliance. This research aims to contribute a timely proof-of-concept.
  - **Feasibility:** The choice of Sentinel-2 data (freely and globally available), a U-Net based architecture (which is powerful yet more manageable in terms of complexity and data requirements compared to very large Transformers), and a focused scope (coffee as the target commodity, selected representative regions in Indonesia) makes the proposed research technically feasible to execute within the 6-month timeframe of a Master's thesis. The strategy to primarily leverage existing ground-truth datasets (with adaptation/validation) or conduct very limited and targeted new validation data collection further enhances this feasibility.
  - **Contribution:** This research is expected to make several contributions:
    - It will provide a validated AI methodology specifically designed for a critical EUDR compliance task in Indonesia, offering a potential pathway

for operationalization.

- It will generate empirical insights into the performance and effectiveness of Siamese U-Net architectures augmented with attention mechanisms for this particular application (fine-grained change detection and segmentation in complex agroforestry landscapes using medium-resolution imagery).
- It may result in the creation of a small but valuable, curated dataset (if new validation data is generated and can be ethically shared) for a specific Indonesian coffee-growing region, which could benefit future research.
- The findings can inform policy discussions and practical efforts regarding the use of AI for environmental regulation, sustainable agriculture, and smallholder inclusion in Indonesia and potentially other commodity-producing countries facing similar challenges.

3. Method

This chapter details the systematic approach that will be undertaken to achieve the research objectives. It outlines the research design, data collection and pre-processing procedures, the proposed AI model architecture and training strategy, and the methods for data analysis and performance evaluation.

3.1. Research Design

This research will adopt an **experimental quantitative research design**. The core of the methodology involves the design and development of a novel Artificial Intelligence (AI) model, its training on a specifically curated geospatial dataset, and a rigorous experimental evaluation of its performance against defined quantitative metrics. The performance of the proposed model will also be compared to a baseline model to assess the impact of the architectural innovations. This approach aligns with common practices in AI and remote sensing research, where empirical evidence is generated through controlled experiments to validate the efficacy of new algorithms or techniques.

Furthermore, the research incorporates elements of **design science research (DSR)**. DSR aims to create and evaluate IT artifacts (in this case, the AI model and the associated workflow) intended to solve identified organizational or societal problems [Hevner et al., 2004]. The problem addressed here is the challenge faced by Indonesian stakeholders in complying with the EUDR, and the proposed AI model is the artifact designed to offer a partial solution.

The research will be executed in a phased manner to ensure systematic progress and feasibility within the 6-month Master's thesis timeframe:

- **Phase 1 (Month 1): In-depth Literature Review Finalization & Data Acquisition Strategy.**
  - Tasks: Complete the detailed review of relevant academic literature and existing solutions. Identify and finalize the specific study regions in Indonesia (e.g., one or two major coffee-producing districts known for significant smallholder presence, such as Gayo in Aceh, a region in Lampung, or Toraja in Sulawesi, based on preliminary assessment of data accessibility, cloud cover conditions, and representativeness of coffee agroforestry systems). Secure access to Sentinel-2 imagery archives for these regions covering relevant time periods (e.g., 2019-2020 for pre-EUDR baseline, and 2021-2024 for post-EUDR monitoring). Finalize the strategy for obtaining ground-truth data: confirm the suitability and accessibility of existing datasets (e.g., adapting AI4SmallFarms [89] if applicable to coffee, or other local/regional datasets) or delineate a clear plan for limited manual annotation of VHR imagery for validation purposes. Develop a detailed data pre-processing pipeline.
  - Deliverable: Finalized literature review, defined study areas, confirmed data access plan, initial pre-processing script outlines.
- **Phase 2 (Month 2): Data Pre-processing & Dataset Curation.**
  - Tasks: Execute the pre-processing of all acquired Sentinel-2 imagery, including atmospheric correction (if L1C data is used), comprehensive cloud and cloud shadow masking. Prepare and process ground-truth data: reformat vector data, re-project to a common CRS, rasterize labels, and generate corresponding image patches and label masks for both deforestation events and coffee plot boundaries. Carefully split the curated dataset into training, validation, and testing sets, ensuring spatial and, where possible, temporal independence to prevent data leakage and biased evaluation.
  - Deliverable: Fully pre-processed image dataset, curated ground-truth labels, and clearly defined training, validation, and test splits.
- **Phase 3 (Month 3): Model Design & Initial Implementation.**
  - Tasks: Implement the proposed Siamese U-Net architecture with integrated attention mechanisms using a deep learning framework such as Python with TensorFlow/Keras or PyTorch. Implement a baseline U-Net model (or a standard change detection algorithm) for comparative performance evaluation. Conduct initial training runs on a small subset of the data to debug the code, verify the data loading pipeline, and ensure the basic model architecture functions as expected.

    - Deliverable: Working code for the proposed and baseline models, successful initial training runs on sample data.
- **Phase 4 (Month 4): Model Training & Hyperparameter Tuning.**
    - Tasks: Conduct full-scale training of the proposed Siamese U-Net with attention and the baseline model(s) on the entire training dataset. Systematically perform hyperparameter tuning (e.g., learning rate, batch size, optimizer parameters, specifics of the attention module configuration, loss function weights) using the validation set to optimize model performance.
    - Deliverable: Trained models with optimized hyperparameters, logs of training progress and validation performance.
- **Phase 5 (Month 5): Model Evaluation & Results Analysis.**
    - Tasks: Evaluate the performance of the finally trained models on the held-out test set using the pre-defined quantitative performance metrics (F1-score, IoU, precision, recall, etc.). Conduct a thorough analysis of the results, including statistical comparisons if applicable. Identify patterns in model errors through qualitative analysis of prediction maps. Assess the model's strengths, weaknesses, and potential utility in the context of EUDR compliance. Visualize key results using maps, charts, and tables.
    - Deliverable: Comprehensive set of evaluation results, comparative analysis, visualizations, and initial interpretation of findings.
- **Phase 6 (Month 6): Thesis Writing, Discussion & Finalization.**
    - Tasks: Complete the writing of all chapters of the Master's thesis, with a particular focus on detailing the methodology, presenting the results, and elaborating on the discussion, conclusions, and recommendations for future work. Incorporate feedback from supervisors and peers. Ensure the thesis document is polished, coherent, and meets all academic standards.
    - Deliverable: Final draft of the Master's thesis.

AI model development is an inherently iterative process.[115] The outlined phases, particularly data pre-processing, model design, training, and hyperparameter tuning, will likely involve cycles of refinement. For instance, initial pre-processing efforts might reveal unforeseen data quality issues that necessitate adjustments to the pipeline.[116] Early model training experiments may indicate the need for modifications to the architecture or a different set of hyperparameters.[53] The 6-month plan, therefore, incorporates an understanding that flexibility and iterative improvement within each phase are crucial for successful AI research. This practical reality means that while the overall structure is sequential, activities within each month may loop back as new findings emerge or challenges are encountered.

3.2. Data Collection

The success of this research hinges on the acquisition and careful curation of appropriate geospatial data. The data collection strategy is designed to be feasible within the thesis timeframe, prioritizing freely available satellite imagery and leveraging existing ground-truth resources where possible.

- Study Area Selection:
  The selection of study areas within Indonesia will be guided by several criteria:
  1. **Prominence of Coffee Cultivation:** Regions known for significant coffee production, particularly by smallholder farmers.
  2. **Presence of Agroforestry Systems:** Areas where coffee is commonly grown in agroforestry or mixed cultivation systems, as these present greater challenges for remote sensing.
  3. **Data Availability:** Consistent availability of relatively cloud-free Sentinel-2 imagery for the required time periods (spanning before and after the EUDR cut-off date of December 31, 2020, and extending to more recent years for monitoring).
  4. **Existing Information:** Availability of some ancillary geospatial data (e.g., coarse land cover maps, administrative boundaries) or accessibility for limited remote validation (e.g., via VHR imagery on platforms like Google Earth).
     Potential candidate regions include, but are not limited to, districts within Aceh (e.g., Gayo highlands), North Sumatra (e.g., Mandailing Natal, Lintong), Lampung, West Java (e.g., Priangan), or South Sulawesi (e.g., Toraja). The final selection of one or two specific districts will be made during Phase 1 based on a detailed assessment of these criteria.
- **Primary Data Sources:**
  - **Satellite Imagery: Sentinel-2 MSI Data**
    - **Source:** Data will be accessed primarily through the Google Earth Engine (GEE) platform [67], which provides analysis-ready data and efficient processing capabilities. Alternatively, data can be sourced from the Copernicus Open Access Hub.[65]
    - **Products:** The Sentinel-2 Level-2A (Surface Reflectance) product will be prioritized as it is atmospherically corrected.[65] If consistent L2A data is unavailable for certain periods/areas, Level-1C (Top-of-Atmosphere reflectance) data will be used and atmospherically corrected during pre-processing. The harmonized Sentinel-2 collections on GEE (e.g., COPERNICUS/S2_SR_HARMONIZED) will be preferred to ensure temporal consistency.[68]
    - **Spectral Bands:** Bands with 10m spatial resolution (Blue - B2, Green - B3,

Red - B4, Near-Infrared - B8) and 20m spatial resolution (Red-Edge bands - B5, B6, B7, B8A; Shortwave Infrared bands - B11, B12) will be the primary inputs for the AI model, as these are most relevant for vegetation analysis and LULC discrimination.[68]
        - **Temporal Resolution:** Imagery will be selected for at least two key time periods: a baseline period prior to the EUDR cut-off date (e.g., imagery from 2019-2020) and one or more periods after the cut-off date (e.g., 2021-2022, 2023-2024) to enable change detection for deforestation. To mitigate cloud cover issues, multiple images within each period may be used to create cloud-free or least-cloudy temporal composites (e.g., median composites).
- **Ancillary Geospatial Data:**
    - **Digital Elevation Model (DEM):** Data from the Shuttle Radar Topography Mission (SRTM) at 30m resolution [43] or a national DEM from Indonesia's INA-Geoportal [97] (if accessible and of suitable quality) will be used. DEM-derived features like slope and aspect can be valuable ancillary inputs for LULC classification and deforestation modeling.
    - **Existing Land Cover Maps:** National or global land cover products (e.g., from Indonesia's Ministry of Environment and Forestry (MoEF), LAPAN/BRIN [99], ESA WorldCover [56], or Google's Dynamic World [39]) will be consulted for contextual information, initial area stratification, or as a reference, although the AI model will perform its own detailed LULC classification relevant to coffee plots.
    - **Administrative Boundaries:** Official administrative boundary data for Indonesia (provinces, districts) will be used to define the precise extent of the study areas and for aggregating and reporting results.
    - **Peatland Maps:** Maps indicating the distribution of peatlands, sourced from the Indonesian Ministry ofAgriculture [92] or other reputable sources, will be used. Deforestation on peatlands is a particularly critical environmental concern and is relevant to EUDR's legality requirements.
- Ground-Truth Data Strategy:
  The availability of accurate ground-truth data is paramount for training and validating the supervised deep learning model. Given the 6-month timeframe, a pragmatic strategy is essential:
    - **Option A (Primary Strategy): Leveraging and Adapting Existing Datasets:**
        - A thorough search will be conducted for existing, publicly available, or research-accessible plot-level datasets for coffee in Indonesia or similar agroforestry systems in Southeast Asia. The **AI4SmallFarms dataset** [89], though focused on Vietnam and Cambodia and covering mixed crops, will be closely examined. Its methodology for delineating smallholder plots

from Sentinel-2 imagery and the structure of its labels are highly relevant. If adaptable (e.g., by identifying coffee plots within it through ancillary data or its attributes, or by using its annotation guidelines as a template), it could provide a valuable starting point, especially for pre-training or transfer learning.

- **Option B (Secondary/Validation Strategy): Limited Manual Annotation of VHR Imagery:**
  - To create a high-quality, independent test/validation set specifically for Indonesian coffee agroforestry systems within the selected study areas, limited manual digitization will be performed. This will involve:
    - Acquiring freely available Very High-Resolution (VHR) satellite imagery (e.g., from Google Earth, Bing Maps, or ESRI's World Imagery basemap) for a small number of representative sub-regions within the study areas, focusing on recent years.
    - Manually digitizing the boundaries of a statistically relevant sample of (a) smallholder coffee plots and (b) recent deforestation patches (post-2020) within these VHR images. The methodology for this will draw inspiration from established practices in ground-truth collection for remote sensing.[91]
    - Visual interpretation will focus on identifying characteristic features of smallholder coffee agroforestry (e.g., irregular plot shapes, mixed canopy textures, proximity to settlements, distinct planting patterns if visible) versus other LULC types like dense forest, other crops, or built-up areas.
- **Deforestation Ground Truth:** For labeling deforestation events, historical deforestation can be initially identified by cross-referencing with existing alert systems like Global Forest Watch GLAD alerts [38] or by adapting methodologies used by INPE in Brazil.[104] These candidate areas will then be visually confirmed and precisely delineated using the temporal sequence of Sentinel-2 imagery, supplemented by VHR imagery where available, to ensure changes correspond to actual deforestation after December 31, 2020.
- **Ethical Considerations for Data:** All geospatial data representing farm locations or potentially identifiable information will be anonymized and aggregated where necessary to protect privacy. This research will primarily rely on publicly available remote sensing data and ethically sourced validation points. Any consideration of participatory data collection (which is unlikely given the scope) would strictly adhere to Free, Prior, and Informed Consent (FPIC) principles.[5]

- **Table 3.1: Characteristics of Proposed Datasets for Model Training and**

Validation

Data Source	Data Type	Resolution/Scale	Temporal Coverage	Geographic Focus in Indonesia (Selected Study Region)	Anticipated Volume (Approx.)	Role in Research
Sentinel-2 (via GEE/Copernicus Hub)	Multispectral Satellite Imagery (L2A/L1C)	10m, 20m (for relevant bands)	2019 – 2024 (pre- & post-EUDR cutoff)	1-2 selected coffee-producing districts	50-100 scenes per district (depending on cloud cover & composites)	Primary input for AI model training, validation, and testing
Google Earth / Bing Maps / ESRI Basemaps	VHR Optical Satellite Imagery	<1m - 2.5m (variable)	Recent years (for validation)	Small sub-regions within study areas	~20-30 sample areas for manual annotation	Creation of high-quality validation/test set for plots & deforestation
AI4SmallFarms (adapted/referenced) [89]	Vector Polygons (plot boundaries), S2 imagery	Polygons derived from 10m S2	2021 (original dataset)	Vietnam/Cambodia (methodology reference)	N/A for direct use; methodology and potential transfer learning	Methodological reference, potential for transfer learning if applicable
SRTM DEM / National DEM	Raster (Elevation)	30m (SRTM) or better	Static	Selected study regions	Coverage for study regions	Ancillary input feature for AI model

Existing LULC Maps (Govt./Global)	Raster/Vector (Land Cover Classes)	Variable (e.g., 30m, 10m)	Various years	National/Regional	Coverage for study regions	Contextual information, initial stratification, reference
GFW Deforestation Alerts [38]	Vector Points/Polygons (Deforestation alerts)	30m (Landsat-based)	Near real-time, historical	Selected study regions	Alerts for study regions	Guidance for identifying candidate deforestation areas for labeling
Manually Annotated Validation Set (from VHR)	Vector Polygons (coffee plots, deforestation)	Derived from VHR (<2.5m)	Post-2020 (for deforestation)	Small sub-regions within study areas	~200-300 coffee plot polygons, ~50-100 deforestation polygons	Independent Test/Validation Set

This table clearly outlines all anticipated data sources, their specifications, and their intended use, demonstrating foresight in data acquisition and management. This is critical for assessing the feasibility of the data collection plan and the robustness of the planned experiments.

The creation of extensive, high-quality ground-truth datasets is often the most significant bottleneck in supervised AI for remote sensing applications.[88] For a 6-month Master's thesis, it is impractical to undertake large-scale de-novo field mapping or manual digitization campaigns. Therefore, the strategy of prioritizing the leveraging and careful adaptation of existing relevant datasets [89], or focusing new data collection efforts *only* on creating small, diverse, and high-quality validation sets using readily available VHR imagery [94], is a more realistic and feasible approach. The core AI model development can proceed with potentially less-than-perfect but available data for the bulk of training, with rigorous and unbiased validation performed on a smaller, more meticulously curated set.

3.3. Data Pre-processing

Thorough and appropriate pre-processing of satellite imagery and ancillary data is crucial for ensuring the quality of inputs to the AI model and, consequently, the reliability of its outputs.

- **Sentinel-2 Imagery Pre-processing:**
  - **Atmospheric Correction:** If Level-1C (Top-of-Atmosphere reflectance) Sentinel-2 data is used, it will be corrected to Level-2A (Surface Reflectance) to minimize the effects of atmospheric scattering and absorption. This can be achieved using established algorithms like Sen2Cor [65] or MAJA [65], often implemented within software like ESA's SNAP toolbox. However, the preferred approach will be to directly utilize the Sentinel-2 L2A surface reflectance collections available in Google Earth Engine (e.g., COPERNICUS/S2_SR_HARMONIZED [68]), as GEE provides analysis-ready data, streamlining this step considerably. The Sen2Chain Python tool also offers capabilities for automating this processing chain if needed.[65]
  - **Cloud and Cloud Shadow Masking:** This is a critical step, especially for optical imagery in tropical regions like Indonesia, which are frequently affected by cloud cover.[65] A robust cloud and cloud shadow masking strategy will be implemented. This will involve utilizing the Sentinel-2 cloud probability layer (derived from the s2cloudless algorithm or similar) and the scene classification layer (SCL) or quality assessment bands (e.g., QA60) provided with Sentinel-2 data. Google Earth Engine offers built-in functionalities and examples for effective cloud masking.[69] The Sen2Chain tool also provides several customizable cloud masking options based on Sen2Cor outputs.[65] The chosen method will aim to accurately identify and mask out pixels contaminated by clouds and their shadows, potentially combining thresholding on cloud probability with morphological filtering to refine the masks.
  - **Mosaicking/Compositing:** To create cloud-free or least-cloudy images for the selected time periods (pre- and post-EUDR cut-off), temporal compositing techniques will be employed. This typically involves selecting the "best" pixel from a stack of images acquired over a defined period (e.g., monthly, seasonally, or annually). Median or mean compositing methods, available in GEE, are commonly used to reduce noise and fill data gaps caused by clouds or shadows.[69]
  - **Co-registration:** Ensuring precise geometric alignment between images

acquired at different times is essential for accurate change detection. Sentinel-2 data provided by ESA and available on GEE is generally well co-registered. This aspect will be verified, and if minor misalignments are detected, image-to-image registration techniques might be explored, though this is less likely to be a major issue with GEE collections.

- **Ancillary Data Pre-processing:**
  - All ancillary vector data (e.g., administrative boundaries, ground-truth polygons) and raster data (e.g., DEM, existing LULC maps) will be re-projected to a common Coordinate Reference System (CRS), typically the UTM zone corresponding to the study area, to ensure spatial consistency.
  - Raster data like DEMs or existing LULC maps will be resampled (e.g., using nearest neighbor or bilinear interpolation, as appropriate) to match the 10m spatial resolution of the primary Sentinel-2 bands used in the analysis.
- Feature Engineering:
  To enhance the discriminative power of the input data for the AI model, several features will be engineered from the Sentinel-2 bands:
  - **Vegetation Indices (VIs):** A set of relevant VIs will be calculated. These include, but are not limited to:
    - Normalized Difference Vegetation Index (NDVI): A common indicator of vegetation greenness and health.
    - Enhanced Vegetation Index (EVI): Similar to NDVI but designed to be more sensitive in high biomass regions and reduce atmospheric influences.
    - Soil Adjusted Vegetation Index (SAVI): Accounts for soil brightness effects.
    - Normalized Burn Ratio (NBR): Useful for identifying burned areas, which can be a precursor or indicator of deforestation. These indices can be computed using standard formulas within GEE's expression() function or using Python libraries like rasterio for local processing.[67] The selection of specific VIs will be guided by literature on coffee crop and forest monitoring.
  - **Texture Features (Exploratory):** If time permits and initial results suggest benefit, texture features derived from Grey Level Co-occurrence Matrix (GLCM) analysis (e.g., contrast, dissimilarity, homogeneity, entropy) could be calculated from the NIR band or a panchromatic-equivalent band. Texture can sometimes help distinguish between different vegetation types with similar spectral responses, such as coffee agroforestry and natural secondary forest.
  - **Band Stacking:** The selected original Sentinel-2 spectral bands, calculated VIs, and potentially DEM-derived features (e.g., elevation, slope) will be stacked together to form multi-channel input image patches for the AI model.
- **Data Formatting for Model Input:**

    - The pre-processed and feature-engineered satellite data will be tiled into smaller image patches (e.g., 128x128 pixels or 256x256 pixels). The optimal patch size will be determined based on the scale of the target features (coffee plots, deforestation areas) and computational constraints.
    - Corresponding label masks will be created for each image patch:
        - For **deforestation detection**, binary masks will indicate pixels that have undergone deforestation (change from forest to non-forest after Dec 31, 2020) versus pixels that have not changed or represent other LULC transitions.
        - For **plot delineation**, binary masks will indicate pixels belonging to coffee plots versus non-plot areas. If finer distinctions are feasible (e.g., coffee vs. other specific crops vs. forest), multi-class masks could be considered, though binary is the primary approach.
    - The dataset of image patches and their corresponding label masks will be carefully divided into training, validation, and testing sets. A common split is 70% for training, 15% for validation (used for hyperparameter tuning and monitoring training progress), and 15% for final, unbiased testing. Stratified sampling based on class prevalence (e.g., proportion of deforestation or coffee plot pixels) will be employed if necessary to ensure that all classes are adequately represented in each set, particularly if there is significant class imbalance.
- Tools for Pre-processing:
  The primary tool for accessing and pre-processing large volumes of Sentinel-2 data will be Google Earth Engine (GEE), accessed via its Python API.67 GEE provides powerful capabilities for cloud masking, mosaicking, and index calculation at scale. For localized processing, patch generation, and management of ground-truth data, Python will be used with libraries such as Rasterio (for raster data I/O and manipulation), GeoPandas (for vector data handling), NumPy (for numerical operations), and Scikit-image (for image processing tasks).

Given Indonesia's tropical environment, characterized by persistent and extensive cloud cover [65], the effectiveness of the cloud masking strategy is not merely a routine pre-processing step but a critical determinant of the project's success. Inaccurate or incomplete cloud masking will introduce substantial noise and errors into the training data, inevitably leading to poor model performance and unreliable results. Therefore, dedicated effort will be invested in selecting, implementing, and potentially comparing different robust cloud masking approaches (e.g., leveraging the s2cloudless dataset in GEE [69], or exploring advanced options available through tools like Sen2Chain if GEE methods prove insufficient for specific challenging areas [65]). This might involve iterative refinement of masking parameters based on visual inspection of sample imagery from the study regions.

3.4. Data Processing (Proposed AI Model/Technique): Siamese U-Net with Attention

The core of this research is the design and application of a specialized deep learning model for the dual tasks of deforestation detection and coffee plot delineation. A Siamese U-Net architecture incorporating attention mechanisms has been chosen for its theoretical suitability to these tasks and its feasibility within the project's scope.

- **Rationale for Chosen Technique:**
    - **U-Net Architecture:** The U-Net is a fully convolutional neural network renowned for its effectiveness in semantic segmentation tasks, particularly in biomedical and remote sensing image analysis.[48] Its characteristic encoder-decoder structure with skip connections allows it to capture both contextual information (through the encoder's downsampling path) and precise localization details (by forwarding high-resolution features from the encoder to the decoder via skip connections). This makes it adept at delineating boundaries, which is crucial for both plot mapping and identifying deforestation patches. Its relative efficiency and ability to train well with moderately sized datasets make it a feasible choice for a 6-month Master's thesis.
    - **Siamese Network Structure:** For the task of change detection (specifically, identifying deforestation that has occurred after a defined cut-off date), a Siamese architecture is highly appropriate.[48] Siamese networks employ two identical (weight-sharing) subnetworks to process two different input images (e.g., satellite images from before and after the EUDR's December 31, 2020, cut-off date). By comparing the feature representations extracted by these parallel encoders, the network can learn to effectively identify and localize areas of change. This directly addresses the EUDR's requirement to pinpoint deforestation events relative to a specific temporal baseline.
    - **Attention Mechanisms:** To enhance the model's ability to discern subtle features and focus on the most informative parts of the input imagery, attention mechanisms will be integrated.[48] In complex agroforestry landscapes, smallholder coffee plots can be spectrally similar to other vegetation types, and early-stage deforestation can be subtle. Attention mechanisms (such as spatial attention, channel attention, or self-attention modules integrated into the decoder or skip connections) can help the model to adaptively weigh the importance of different features or spatial locations,

thereby improving its sensitivity to small or irregularly shaped coffee plots and fine-scale deforestation signals. This is hypothesized to yield better performance compared to a vanilla U-Net.

- Proposed Model Architecture (Detailed Description):
  The proposed model will be a Siamese U-Net with attention. A conceptual diagram of the architecture will be provided in the final thesis.
  - **Input:**
    - For **deforestation detection**: The model will take pairs of co-registered image patches (e.g., 128x128 or 256x256 pixels, composed of selected Sentinel-2 bands and derived VIs) from two time points: T1 (representing the pre-EUDR cut-off period, e.g., 2020) and T2 (representing the post-EUDR cut-off period, e.g., 2021-2024).
    - For **coffee plot delineation**: The primary input will be a single image patch from T2 (the more recent period). The Siamese structure can be adapted for this by either feeding the T2 patch to both encoder branches (allowing the network to learn robust features even in a single-image context) or by using the T1 patch as contextual information in one branch if it proves beneficial. Alternatively, a separate, standard U-Net with attention might be trained specifically for delineation if the Siamese structure is not optimal for single-image segmentation. This will be explored during initial experimentation.
  - **Siamese Encoder:** The encoder will consist of two identical branches, each being a U-Net-style contracting path. This path will typically comprise multiple convolutional blocks (e.g., two 3x3 convolutions followed by ReLU activation and batch normalization) and max-pooling layers for downsampling. A pre-trained backbone (e.g., a lightweight ResNet variant) could be considered for the encoder blocks to leverage features learned from larger datasets, though training from scratch is also an option.
  - **Feature Fusion/Comparison (for Change Detection):** At the bottleneck of the U-Net (the lowest resolution feature maps) and potentially at intermediate levels via skip connections, features extracted by the two Siamese encoder branches will be combined to highlight differences. Common fusion methods include element-wise subtraction or concatenation of the feature maps from T1 and T2.
  - **Decoder with Attention:** The decoder will follow a U-Net-style expansive path, using up-convolution (transposed convolution) or upsampling layers (e.g., bilinear upsampling followed by convolution) to gradually increase the spatial resolution of the feature maps back to the original input patch size.
    - **Attention Modules:** Attention blocks will be integrated within the decoder.

These could be:

        - **Spatial Attention Modules:** To help the network focus on salient spatial regions.
        - **Channel Attention Modules (e.g., Squeeze-and-Excitation blocks):** To adaptively recalibrate channel-wise feature responses.
        - Self-attention layers (if computationally feasible) at certain stages of the decoder.
    - **Skip Connections:** Features from corresponding levels of the encoder path(s) will be concatenated with the upsampled features in the decoder, allowing the network to reuse fine-grained details for precise localization. Attention mechanisms might also be applied to these skip connections.
  - **Output Layer(s):**
    - For **deforestation detection**: A final 1x1 convolutional layer followed by a sigmoid activation function to produce a pixel-wise probability map, where each pixel value represents the likelihood of deforestation (change from forest to non-forest).
    - For **coffee plot delineation**: A final 1x1 convolutional layer followed by a sigmoid activation function for binary classification (coffee plot vs. non-plot background). If distinguishing multiple LULC types becomes a secondary objective, a softmax activation would be used for multi-class segmentation.
  - **Model Implementation:** The model will be implemented in Python using either TensorFlow (with the Keras API) or PyTorch, leveraging their extensive libraries for deep learning.
- **Training Strategy:**
  - **Dataset Split:** The curated dataset of image patches and corresponding labels will be divided into three sets: typically 70% for training, 15% for validation (used for hyperparameter tuning and monitoring for overfitting), and 15% for final, unbiased testing. Stratified sampling will be considered to ensure that all classes (especially minority classes like deforestation) are adequately represented in each split.
  - **Loss Function:**
    - For the **segmentation tasks** (plot delineation and deforestation as a binary segmentation problem), a compound loss function combining Dice Loss (or Tversky Loss for better handling of class imbalance) and Binary Cross-Entropy (BCE) Loss will likely be used. Dice Loss is effective for improving IoU, while BCE helps with pixel-wise classification accuracy.
    - For **change detection**, if framed as a direct binary classification of change, BCE Loss or Focal Loss (which down-weights well-classified

examples and focuses on hard, misclassified ones, useful if change pixels are rare) will be considered.
  - **Optimizer:** The Adam or AdamW (Adam with decoupled weight decay) [Loshchilov and Hutter, 2019] optimizer will be used due to their adaptive learning rates and good general performance. An appropriate learning rate schedule, such as cosine annealing or "reduce learning rate on plateau" (monitoring validation loss), will be employed to facilitate convergence and avoid local minima.
  - **Hyperparameter Tuning:** Key hyperparameters—including learning rate, batch size, number of training epochs, architectural details of the attention modules, and weights for compound loss functions—will be tuned systematically. This may involve manual experimentation or automated methods like grid search or random search, with performance evaluated on the validation set.
  - **Data Augmentation:** To increase the diversity of the training set and improve model generalization, especially if the initial curated dataset is limited, standard geometric data augmentation techniques will be applied on-the-fly during training. These include random flips (horizontal and vertical), rotations, and potentially slight scaling or translations.[39] Photometric augmentations (e.g., adjustments to brightness, contrast, saturation) might also be explored, but with caution, as they could interfere with the subtle spectral signals needed for change detection. Mosaic augmentation, mentioned in [91] for LULC, could also be considered if appropriate.
  - **Transfer Learning (Consideration):** If a suitable pre-trained U-Net or encoder backbone (e.g., ResNet pre-trained on ImageNet or a large remote sensing dataset like those used for AI4SmallFarms [89]) is available, transfer learning will be explored. This involves initializing parts of the model with pre-trained weights and then fine-tuning the entire model or specific layers on the target Indonesian coffee dataset. This can often lead to faster convergence and better performance, especially with limited training data. However, the primary plan is to train the model from scratch on the curated dataset to ensure specificity to the task.
- Baseline Model for Comparison:
  To rigorously evaluate the benefits of the proposed Siamese U-Net with attention, its performance will be compared against a simpler baseline model.
  - For **plot delineation**, a standard U-Net architecture (without the Siamese structure or explicit attention mechanisms) will be trained and evaluated on the same dataset.
  - For **deforestation detection**, a common U-Net based change detection

approach (e.g., by concatenating pre- and post-change images as input to a standard U-Net, or a U-Net that processes subtracted images) will serve as a baseline. This comparative analysis will help quantify the specific contributions of the Siamese architecture and the attention mechanisms to the model's performance.

- Software and Frameworks:
  The project will primarily utilize Python as the programming language. Deep learning model development will be carried out using either TensorFlow (with the Keras API) or PyTorch. Google Earth Engine (via its Python API) will be extensively used for accessing, pre-processing, and analyzing Sentinel-2 satellite data at scale. Other essential Python libraries will include GDAL/Rasterio and GeoPandas for geospatial data manipulation, NumPy for numerical computations, Scikit-learn for performance metrics calculation and data splitting, and Matplotlib/Seaborn/Plotly for data visualization.

The tasks of plot delineation and deforestation detection, while addressed by a dual-output model or closely linked models, are inherently related. A plot of land cannot be classified as deforested unless it was recently forest, and deforestation often occurs to create new agricultural plots. While this thesis proposes to tackle these tasks either sequentially or with a model architecture that handles both, a potential avenue for future research (beyond the 6-month scope) could be the exploration of a true multi-task learning framework. In such a framework, a single, unified model would learn to perform both plot delineation and deforestation detection simultaneously, potentially leveraging shared feature representations to improve the performance and efficiency of both tasks. This is a more advanced concept but represents a logical progression from the work proposed herein.

3.5. Data Analysis (Performance Evaluation)

A rigorous evaluation of the trained models is essential to assess their accuracy, reliability, and potential utility for EUDR compliance. Both quantitative and qualitative analysis methods will be employed.

- Performance Metrics:
  The choice of performance metrics will be tailored to the specific nature of the tasks (binary classification for change detection, semantic segmentation for plot delineation).
  - **For Deforestation Detection (Change Detection - Binary Classification of pixels as 'deforested' vs. 'not deforested'):**
    - **Pixel-level metrics:**

        - **Overall Accuracy (OA):** The proportion of correctly classified pixels. While common, it can be misleading in cases of high class imbalance (e.g., if deforestation is rare).
        - **Precision (Positive Predictive Value):** For the 'deforested' class, this measures the proportion of pixels predicted as deforested that were actually deforested. (Formula: TP/(TP+FP)).[120] High precision minimizes false alarms of deforestation.
        - **Recall (Sensitivity, True Positive Rate):** For the 'deforested' class, this measures the proportion of actual deforested pixels that were correctly identified by the model. (Formula: TP/(TP+FN)).[120] High recall is crucial for EUDR to minimize missed deforestation events.
        - **F1-Score:** The harmonic mean of Precision and Recall, providing a balanced measure, especially useful when classes are imbalanced. (Formula: 2×(Precision×Recall)/(Precision+Recall)).[48]
        - **Specificity (True Negative Rate):** For the 'deforested' class, this measures the proportion of actual non-deforested pixels correctly identified. (Formula: TN/(TN+FP)).
        - **Intersection over Union (IoU) / Jaccard Index:** For the 'deforested' class, this measures the overlap between the predicted deforestation mask and the ground-truth deforestation mask. (Formula: TP/(TP+FP+FN)).[48] It is a standard metric for segmentation quality.
    - **For Coffee Plot Delineation (Semantic Segmentation - Binary Classification of pixels as 'coffee plot' vs. 'non-plot'):**
        - **Pixel-level metrics:** Similar to deforestation detection, OA, Precision, Recall, F1-Score, Specificity, and IoU will be calculated for the 'coffee plot' class.[48]
        - **Object-level metrics (if feasible and ground truth allows):**
            - **PoLiS (Polygon Similarity) metric:** If the ground truth consists of precise vector polygons for coffee plots, the PoLiS metric can be used to evaluate the geometric similarity (positional and shape accuracy) between predicted plot polygons and reference polygons.[89] This requires converting the model's raster output to vector polygons.
    - **Kappa Coefficient (Cohen's Kappa):** This metric measures the agreement between the model's classification and the ground truth, while accounting for the agreement that could occur by chance. It is a robust measure of classification accuracy, especially for LULC mapping.[53]
- **Justification of Metrics:**
    - The **F1-Score and IoU** are particularly important because they provide a balanced assessment of segmentation quality, especially in scenarios with

potential class imbalance (e.g., deforested pixels might be a small fraction of the total area, or coffee plots might cover a smaller area than non-plot land). A high F1-score indicates a good balance between precision (not too many false alarms) and recall (not missing too many actual instances).[120] IoU is a standard and intuitive measure of segmentation overlap.
  - **Precision and Recall** are individually crucial for understanding specific aspects of model performance in the EUDR context.[120] For deforestation detection, high recall is paramount to ensure that actual instances of deforestation (which would indicate non-compliance) are not missed by the system. For plot delineation, high precision is important to avoid incorrectly identifying areas as coffee plots, which could lead to unnecessary scrutiny or burden on landowners. High recall for plots ensures that most actual coffee plots are identified for monitoring.
  - The choice of these metrics aligns with common best practices in remote sensing image analysis and machine learning evaluation.[48]
- Statistical Significance Testing:
  If multiple models are being compared (e.g., the proposed model versus the baseline U-Net), appropriate statistical tests will be considered to determine if the observed differences in performance metrics are statistically significant. Depending on the experimental setup (e.g., k-fold cross-validation), tests such as McNemar's test (for paired nominal data) or paired t-tests on the performance scores from cross-validation folds might be applicable.
- Qualitative Analysis:
  Beyond quantitative metrics, a qualitative analysis of the model's outputs is essential. This will involve:
  - Visual inspection of the predicted segmentation masks (for both deforestation and plots) overlaid on the original satellite imagery and compared against the ground-truth labels for representative areas within the test set.
  - Identification of common error types, such as:
    - Boundary errors (e.g., over-segmentation or under-segmentation of plot edges).
    - Missed detections of small or isolated coffee plots or deforestation patches.
    - False positives (e.g., misclassifying other vegetation types as coffee, or stable forest as deforested).
    - Areas of confusion (e.g., difficulty distinguishing shaded coffee from sparse forest or other agroforestry systems).
  - Analysis of model performance in challenging conditions (e.g., areas with residual cloud/haze if masking was imperfect, areas with highly

heterogeneous landscapes).

- **Table 3.2: Selected Performance Metrics and Justification**

Metric	**Formula/Definition**	**Relevance to Deforestation Detection Task (EUDR Context)**	**Relevance to Plot Delineation Task (EUDR Context)**
Precision	TP/(TP+FP)	High value minimizes false alarms of deforestation, reducing unnecessary investigations.	High value minimizes misclassifying non-coffee areas as coffee plots, avoiding incorrect attribution of production.
Recall (Sensitivity)	TP/(TP+FN)	**Crucial:** High value ensures most actual deforestation events are detected, minimizing risk of overlooking non-compliance.	High value ensures most actual coffee plots are identified for monitoring and traceability.
F1-Score	2×(Precision×Recall)/(Precision+Recall)	Balanced measure of precision and recall; good overall indicator of deforestation detection performance.	Balanced measure of precision and recall; good overall indicator of plot delineation accuracy, especially with imbalanced plot sizes/numbers.
IoU (Jaccard Index)	TP/(TP+FP+FN)	Measures the spatial overlap accuracy of detected deforestation areas; robust for segmentation quality.	Measures the spatial overlap accuracy of delineated plots; key for assessing boundary accuracy.
Overall Accuracy	(TP+TN)/(TP+TN+FP+FN)	General measure, but can be misleading if deforestation class is highly imbalanced.	General measure, but less informative than IoU or F1 for segmentation quality if plot class is imbalanced.

Kappa Coefficient	Accounts for chance agreement	Robust measure of classification agreement beyond chance for deforestation mapping.	Robust measure of classification agreement beyond chance for coffee plot mapping.
PoLiS (Object-level)	Compares geometric similarity of predicted vs. reference polygons	N/A (primarily for object delineation)	If feasible, provides a strong measure of how well the shape and location of individual plots are captured.

This table explicitly links the chosen evaluation metrics to the research objectives and the practical implications for EUDR compliance. It demonstrates a clear understanding of why certain metrics are important (e.g., high recall for deforestation to avoid missing non-compliant areas). This demonstrates rigor in the evaluation plan.

3.6. Data Visualisation

Effective visualization of data and results is crucial for interpreting the model's performance and communicating the research findings. The following visualizations will be generated:

- **Input Data Visualization:** Examples of pre-processed Sentinel-2 image patches (e.g., true-color composites, false-color composites highlighting vegetation) and corresponding ground-truth masks for both deforestation and coffee plots will be presented to illustrate the nature of the input data.
- **Model Output Visualization:**
  - Maps displaying the original Sentinel-2 imagery, the ground-truth labels (deforestation and plots), and the model's predicted segmentation masks for representative areas within the test set. These will allow for a visual assessment of the model's accuracy and error patterns.
  - Side-by-side comparisons of the outputs from the proposed Siamese U-Net with attention and the baseline U-Net model for the same test areas to visually highlight performance differences.

- **Quantitative Performance Visualization:**
    - Confusion matrices for both the deforestation detection task and the plot delineation task, showing the distribution of true positives, true negatives, false positives, and false negatives.
    - Receiver Operating Characteristic (ROC) curves and their associated Area Under the Curve (AUC) values for the binary classification tasks.
    - Precision-Recall (PR) curves, which are particularly informative for imbalanced datasets.
    - Bar charts or tables summarizing the key quantitative performance metrics (F1-score, IoU, precision, recall, Kappa) for the proposed model and the baseline model, potentially including error bars if cross-validation is performed.
- **Error Analysis Visualization:**
    - Examples of challenging cases where the model performed poorly (e.g., misclassification of small or irregular plots, confusion with similar land cover types, errors in areas with residual cloud effects or complex topography). These visualizations will help in understanding the model's limitations.
    - If specific patterns of error are identified (e.g., related to plot size, shade levels in agroforestry, or proximity to forest edges), these might be visualized through scatter plots or histograms.

These visualizations will be created using Python libraries such as Matplotlib, Seaborn, and potentially interactive plotting libraries like Plotly if suitable for inclusion in the thesis document (e.g., as static exports). Geospatial visualizations (maps) will be generated using libraries like GeoPandas in conjunction with Matplotlib, or by exporting model outputs for display in GIS software like QGIS.

3.7. Data Validity

Ensuring the validity of the research findings is paramount. Several measures will be taken to address potential threats to data validity and model reliability:

- **Cross-Validation (if feasible):** If the size of the curated training dataset permits without making individual folds too small, k-fold cross-validation (e.g., 5-fold or 10-fold) will be employed during the model training and hyperparameter tuning phase. This involves splitting the training data into k subsets, training the model k times (each time using k-1 subsets for training and one for validation), and averaging the performance metrics across the k folds. This provides a more robust estimate of the model's generalization performance on unseen data and helps in mitigating overfitting to a specific train-validation split.

- **Test Set Integrity:** A dedicated, independent test set (e.g., 15% of the total dataset) will be strictly held out and used *only once* for the final evaluation of the trained models (both the proposed model and the baseline). This ensures an unbiased assessment of the model's ability to generalize to new, previously unseen data. The test set will not be used for any part of the training or hyperparameter tuning process.
- **Addressing Potential Biases:**
  - **Data Bias / Class Imbalance:** The distribution of classes in the training data (e.g., proportion of deforested pixels vs. non-deforested pixels; coffee plot pixels vs. background pixels) will be carefully analyzed. If significant class imbalance is detected, which is common in remote sensing applications (e.g., deforestation events are often rare compared to stable areas), appropriate techniques will be employed to mitigate its impact on model training. These may include:
    - Using weighted loss functions that assign higher penalties to errors made on the minority class.
    - Employing oversampling techniques for the minority class (e.g., SMOTE - Synthetic Minority Over-sampling Technique, though with caution for image data) or undersampling the majority class in the training set.
    - Utilizing data augmentation more extensively for the minority class.
  - **Geographic Bias:** If the ground-truth data is predominantly sourced from or validated in a very specific sub-region within the larger study area, this potential geographic bias will be acknowledged as a limitation. The aim during study area selection and ground-truth strategy finalization (Phase 1) will be to ensure that the data is as representative as possible of the coffee agroforestry systems and deforestation patterns within the chosen districts, within the practical constraints of data availability.
  - **Model Bias:** The qualitative error analysis (Section 3.5) will play a crucial role in identifying if the model exhibits systematic biases, such as consistently underperforming on certain types of coffee plots (e.g., very small plots, plots under dense shade canopy, specific intercropping patterns) or particular types of deforestation (e.g., small-scale selective logging vs. clear-cuts). Understanding these biases is key to interpreting the model's real-world applicability.
- **Sensitivity Analysis (Exploratory, if time permits):** If time allows after the primary model evaluation, a brief sensitivity analysis could be conducted. This might involve exploring how the model's performance changes with variations in key input parameters (e.g., the threshold used for cloud masking, the size of the input image patches, or the specific combination of vegetation indices used as

input features). This can provide insights into the model's robustness to variations in data processing choices.

A critical aspect of robust AI research involves not just reporting high accuracy metrics but also conducting a thorough analysis of "failure cases".[115] Understanding *where* and *why* the proposed model makes errors is crucial. For instance, if the model consistently fails to detect small deforestation patches under dense canopy, or if it frequently misclassifies a particular type of coffee agroforestry system as non-coffee land, these are critical findings. Such an analysis, which will be part of the "Results and Discussion" chapter, adds significant depth to the thesis by highlighting the model's current limitations, assessing its true real-world utility for EUDR compliance, and providing clear directions for future research and model improvement.

- **Table 3.3: Provisional 6-Month Research Timeline**

Phase	Month	Key Tasks per Month	Deliverables/Milestones for each Month
1. Foundation & Data Strategy	1	Finalize detailed literature review. Select study regions. Confirm Sentinel-2 data access. Finalize ground-truth data strategy (existing datasets/VHR annotation plan). Develop detailed data pre-processing pipeline.	Completed Lit Review Chapter outline. Defined Study Areas. Data Acquisition Plan. Pre-processing Protocol.
2. Data Acquisition & Pre-processing	2	Acquire/access Sentinel-2 imagery. Execute image pre-processing (atmospheric correction, cloud masking, mosaicking). Prepare/adapt ground-truth data. Create image patches & labels.	Pre-processed Sentinel-2 imagery for study areas. Curated ground-truth dataset. Finalized Train/Validation/Test splits.

		Split data (train/val/test).	
3. Model Design & Initial Implementation	3	Implement Siamese U-Net with Attention. Implement baseline U-Net. Develop data loaders. Conduct initial training runs on subset of data for debugging and pipeline verification.	Working code for proposed & baseline models. Successful test runs on sample data. Refined model architecture diagrams.
4. Model Training & Hyperparameter Tuning	4	Conduct full training of proposed and baseline models. Perform systematic hyperparameter tuning using the validation set. Monitor for overfitting. Document training process.	Optimized trained models (proposed & baseline). Training logs and validation performance curves. Hyperparameter tuning report.
5. Model Evaluation & Results Analysis	5	Evaluate trained models on the held-out test set using all defined metrics. Perform quantitative and qualitative error analysis. Compare model performances. Visualize results. Interpret findings.	Comprehensive evaluation results. Comparative performance analysis. Visualizations (maps, charts). Draft of Results chapter.
6. Thesis Writing, Discussion & Finalization	6	Complete writing of all thesis chapters (Introduction, Literature Review, Methodology, Results, Discussion, Conclusion, Future Work). Incorporate supervisor feedback.	Full first draft of thesis. Revised draft after feedback. Final Thesis Submission.

		Final proofreading and formatting.	

This table is essential for a Master's thesis proposal as it demonstrates project management skills and the feasibility of completing the research within the given timeframe. It provides a clear roadmap for the student and allows the supervisor to track progress, forcing realistic planning of complex tasks.

Contribution to the Field of AI in General and in Government:

This research is poised to make meaningful contributions both to the broader field of Artificial Intelligence and specifically to the application of AI in governmental efforts related to environmental regulation and sustainable agriculture.

- **Contributions to the Field of AI in General:**
    - The study will contribute to the expanding body of research on the application of advanced deep learning architectures, specifically Siamese networks and U-Nets augmented with attention mechanisms, to complex geospatial image analysis tasks. It will provide empirical evidence on the efficacy of these models for fine-grained change detection (deforestation) and semantic segmentation (plot delineation) in challenging, heterogeneous environments characterized by small target features.
    - It will offer insights into the performance of these sophisticated architectures when applied to freely available, medium-resolution Sentinel-2 satellite data for tasks that often push the limits of such imagery (e.g., delineating smallholder plots typically requiring VHR data). This can inform the development of more robust models that can effectively leverage widely accessible data sources.
    - If new, albeit limited, curated datasets for Indonesian coffee plots and associated deforestation are generated as part of the validation process and can be ethically shared (e.g., anonymized plot boundaries and corresponding image patches), this would represent a valuable resource for the research community, fostering further work in this specific domain and geographic region.
    - The comparative analysis against a baseline U-Net will provide quantitative evidence of the benefits (or limitations) of incorporating Siamese structures

and attention mechanisms for these specific tasks, contributing to a better understanding of architectural choices in deep learning for remote sensing.

- **Contributions to AI in Government (specifically for Indonesian EUDR Compliance):**
    - The research will deliver a proof-of-concept for a technically feasible AI-driven tool that can directly support Indonesian government agencies (such as the Ministry of Agriculture, the Ministry of Environment and Forestry, and the newly formed National EUDR Task Force) as well as other key stakeholders (e.g., farmer cooperatives, exporters, certification bodies) in addressing the demanding due diligence requirements of the EUDR.
    - The methodology developed for coffee plot delineation, if successful, can inform and potentially enhance national efforts to update or refine land cover databases and agricultural plot inventories, contributing to more accurate national statistics and resource management.
    - The deforestation detection outputs, particularly when linked to specific plots, can be integrated into or complement existing national forest monitoring systems, such as the planned National Dashboard or independent initiatives like GTID [17], thereby strengthening Indonesia's capacity for environmental oversight and law enforcement related to deforestation.
    - This thesis will serve as a case study on how state-of-the-art AI techniques can be adapted and applied to address specific national regulatory challenges in the agricultural sector. The lessons learned and the methodology developed could potentially be replicated or adapted for other EUDR-affected commodities in Indonesia or for similar regulatory contexts in other countries.
    - By specifically focusing on smallholder coffee plots, the research directly addresses a key concern of the Indonesian government regarding the EUDR's socio-economic impact on vulnerable

Works cited

1. EU Deforestation Regulation: Key Insights for Global Supply Chains - Sedex, accessed May 16, 2025, https://www.sedex.com/blog/eu-deforestation-regulation-key-insights-for-global-supply-chains/
2. 10 key things you STILL need to know about the new EU ..., accessed May 16, 2025, https://www.whitecase.com/insight-alert/10-key-things-you-still-need-know-about-new-eu-deforestation-regulation
3. Deforestation Regulation implementation - European Commission, accessed May 16, 2025,

https://green-forum.ec.europa.eu/deforestation-regulation-implementation_en
4. EUDR for dummies: Your guide to navigating the EU Deforestation Regulation, accessed May 16, 2025, https://www.positiongreen.com/insights/articles/eudr-for-dummies-your-guide-to-navigating-the-eudr/
5. Why Southeast Asia should seize the EUDR compliance opportunity - EY, accessed May 16, 2025, https://www.ey.com/en_id/insights/sustainability/why-southeast-asia-should-seize-the-eudr-compliance-opportunity
6. How to comply with the EUDR: A step-by-step guide - Coolset, accessed May 16, 2025, https://www.coolset.com/academy/how-to-comply-with-the-eudr-guide
7. Indonesia raises concerns over EU deforestation law's impact on smallholders - Mongabay, accessed May 16, 2025, https://news.mongabay.com/2025/04/indonesia-raises-concerns-over-eu-deforestation-laws-impact-on-smallholders/
8. Indonesian palm oil smallholders and the EUDR - Fern.org, accessed May 16, 2025, https://www.fern.org/fileadmin/uploads/fern/Documents/2025/Fern_Indonesian_palm_oil_smallholders_and_the_EUDR_impacts_and_ways_forward.pdf
9. EUDR Compliance · Picterra, accessed May 16, 2025, https://picterra.ch/our_services/eudr-compliance/
10. Press: IntegrityNext Launches Enhanced EUDR Solution, accessed May 16, 2025, https://www.integritynext.com/about-us/press/detail/integritynext-launches-enhanced-eudr-solution-to-automate-due-diligence-and-ensure-supply-chain-transparency
11. Understanding EUDR Penalties: 10 Business Implications of Non-Compliance - LiveEO, accessed May 16, 2025, https://www.live-eo.com/article/eudr-non-compliance-penalties
12. European Union: EU Deforestation Regulation - Revised Implementation Timeline for 2025, accessed May 16, 2025, https://www.fas.usda.gov/data/european-union-eu-deforestation-regulation-revised-implementation-timeline-2025
13. Deforestation Regulation (EUDR): 4 AI Powered Cost-Saving ..., accessed May 16, 2025, https://diss-co.tech/eu-deforestation-regulation-eudr/
14. Indonesia's path to EUDR compliance: Turning challenges into ..., accessed May 16, 2025, https://www.techedt.com/indonesias-path-to-eudr-compliance-turning-challenges-into-opportunities
15. AI and Blockchain Technologies Can Aid Companies in EUDR Compliance - Earth.Org, accessed May 16, 2025, https://earth.org/ai-and-blockchain-technologies-can-aid-companies-in-eudr-compliance/
16. The Impact of the European Union Anti-Deforestation Regulation ..., accessed May 16, 2025, https://esl.ipb.ac.id/the-impact-of-the-european-union-anti-deforestation-regulation-eudr-on-sustainable-agricultural-development-in-indonesia/

17. Indonesia strengthens forest monitoring with new tool to meet EU deforestation law, accessed May 16, 2025, https://news.mongabay.com/2025/04/indonesia-strengthens-forest-monitoring-with-new-tool-to-meet-eu-deforestation-law/
18. Reviews of smallholder challenges and indigenous oil palm production in Indonesia, accessed May 16, 2025, https://efi.int/news/reviews-smallholder-challenges-and-indigenous-oil-palm-production-indonesia-2025-01-20
19. At the End of the Chain: The Implications of the EU Deforestation Regulation on Smallholders in the Indonesian Palm Oil Trade - Knowledge UChicago, accessed May 16, 2025, https://knowledge.uchicago.edu/record/14665
20. Smallholder Inclusion Under the EUDR: Opportunity or Obstacle ..., accessed May 16, 2025, https://farmerline.co/smallholder-inclusion-under-the-eudr-opportunity-or-obstacle/
21. Indonesian palm oil smallholders and the EUDR: Impacts and ways forward - Fern.org, accessed May 16, 2025, https://www.fern.org/publications-insight/indonesian-palm-oil-smallholders-and-the-eudr-impacts-and-ways-forward/
22. Finding a place for smallholder farmers in EU deforestation regulation | SEI, accessed May 16, 2025, https://www.sei.org/publications/smallholder-farmers-eu-deforestation/
23. Minimize EUDR Risk in Deforestation Free Supply Chain - TraceX Technologies, accessed May 16, 2025, https://tracextech.com/minimize-eudr-risk-in-deforestation-free-supply-chain/
24. Indonesia Strengthens Forest Monitoring to Comply with EU Deforestation Law, Rubber Industry Faces Transformation - ECHEMI.com, accessed May 16, 2025, https://www.echemi.com/cms/2361119.html
25. Ground-truthed.id – Kaoem Telapak, accessed May 16, 2025, https://kaoemtelapak.org/ground-truthed-id/
26. GTID: A Digital Monitoring Platform to Strengthen Traceability and ..., accessed May 16, 2025, https://kaoemtelapak.org/gtid-a-digital-monitoring-platform-to-strengthen-traceability-and-environmental-advocacy/
27. accessed January 1, 1970, https://www.mongabay.com/2025/04/indonesia-strengthens-forest-monitoring-with-new-tool-to-meet-eu-deforestation-law/
28. AI + Forest Monitoring: How Coolx Supports EUDR Compliance | Nature Tech Collective, accessed May 16, 2025, https://www.naturetechcollective.org/stories/forest-monitoring-eudr-compliance-coolx
29. AI agricultural compliance: transforming farm tech for smarter regulation - BytePlus, accessed May 16, 2025, https://www.byteplus.com/en/topic/563312
30. Navigating EUDR Landscape: How Real-Time Data Enables ..., accessed May 16, 2025, https://www.cropin.com/blogs/manage-eudr-compliance.html

31. Satellite Monitoring for Deforestation: Tools & Solutions - FlyPix AI, accessed May 16, 2025, https://flypix.ai/blog/satellite-monitoring-deforestation-solutions-sowtware-and-tools/
32. Technologies to Protect Global Forests: The Case of Indonesia - Istituto Affari Internazionali, accessed May 16, 2025, https://www.iai.it/en/pubblicazioni/c05/technologies-protect-global-forests-case-indonesia
33. How geospatial AI can help you comply with EU's deforestation law · Customers Picterra, accessed May 16, 2025, https://picterra.ch/blog/how-geospatial-ai-can-help-you-comply-with-eus-deforestation-law/
34. TradeAware - The End-To-End EUDR Solution I LiveEO, accessed May 16, 2025, https://www.live-eo.com/product/tradeaware
35. Ethical Considerations In Ai Driven Agriculture Expansion - Prism → Sustainability Directory, accessed May 16, 2025, https://prism.sustainability-directory.com/scenario/ethical-considerations-in-ai-driven-agriculture-expansion/
36. (PDF) A comprehensive GeoAI review: Progress, Challenges and Outlooks - ResearchGate, accessed May 16, 2025, https://www.researchgate.net/publication/387105937_A_comprehensive_GeoAI_review_Progress_Challenges_and_Outlooks
37. [2405.20868] Responsible AI for Earth Observation - arXiv, accessed May 16, 2025, https://arxiv.org/abs/2405.20868
38. Data and Methods | Global Forest Review, accessed May 16, 2025, https://gfr.wri.org/data-methods
39. Tracking U.S. Land Cover Changes: A Dataset of Sentinel-2 Imagery and Dynamic World Labels (2016–2024) - MDPI, accessed May 16, 2025, https://www.mdpi.com/2306-5729/10/5/67
40. unfccc.int, accessed May 16, 2025, https://unfccc.int/ttclear/misc_/StaticFiles/gnwoerk_static/AI4climateaction/ea0f2596d93640349b9b65f4a7c7dd24/b47ef0e99cb24e57aa9ea69f0f5d6a71.pdf
41. Remote sensing and machine learning for environmental monitoring: Opportunities and challenges - AI for Good, accessed May 16, 2025, https://aiforgood.itu.int/event/remote-sensing-and-machine-learning-for-environmental-monitoring-opportunities-and-challenges/
42. State of the art in remote sensing monitoring of carbon ... - Frontiers, accessed May 16, 2025, https://www.frontiersin.org/journals/remote-sensing/articles/10.3389/frsen.2025.1532280/full
43. Predictive Modelling of Land Cover Changes in the Greater ... - MDPI, accessed May 16, 2025, https://www.mdpi.com/2072-4292/16/21/4013
44. Transactions on Geoscience & Remote Sensing - GRSS-IEEE, accessed May 16, 2025, https://www.grss-ieee.org/publications/transactions-on-geoscience-remote-sen

sing/
45. Remote Sensing Data Quality in the Era of AI - ResearchGate, accessed May 16, 2025, https://www.researchgate.net/publication/385524547_Remote_Sensing_Data_Quality_in_the_Era_of_AI
46. Artificial intelligence in environmental monitoring: in-depth analysis, accessed May 16, 2025, https://d-nb.info/1355324793/34
47. Artificial intelligence and IoT driven technologies for environmental pollution monitoring and management - ResearchGate, accessed May 16, 2025, https://www.researchgate.net/publication/378315907_Artificial_intelligence_and_IoT_driven_technologies_for_environmental_pollution_monitoring_and_management
48. (PDF) A Full-Scale Connected CNN–Transformer Network for ..., accessed May 16, 2025, https://www.researchgate.net/publication/385491574_A_Full-Scale_Connected_CNN-Transformer_Network_for_Remote_Sensing_Image_Change_Detection
49. Advancements in Visual Language Models for Remote Sensing: Datasets, Capabilities, and Enhancement Techniques - arXiv, accessed May 16, 2025, https://arxiv.org/html/2410.17283v3
50. Comparative and Interpretative Analysis of CNN and Transformer Models in Predicting Wildfire Spread Using Remote Sensing Data | Request PDF - ResearchGate, accessed May 16, 2025, https://www.researchgate.net/publication/390643841_Comparative_and_Interpretative_Analysis_of_CNN_and_Transformer_Models_in_Predicting_Wildfire_Spread_Using_Remote_Sensing_Data
51. Change Detection for Forest Ecosystems Using Remote Sensing Images with Siamese Attention U-Net, accessed May 16, 2025, https://d-nb.info/1346891826/34
52. Land cover mapping with Sentinel-2 imagery using deep learning semantic segmentation models - CEUR-WS.org, accessed May 16, 2025, https://ceur-ws.org/Vol-3909/Paper_1.pdf
53. (PDF) Optimised U-Net for Land Use–Land Cover Classification ..., accessed May 16, 2025, https://www.researchgate.net/publication/368482330_Optimised_U-Net_for_Land_Use-Land_Cover_Classification_Using_Aerial_Photography
54. A Review of CNN Applications in Smart Agriculture Using Multimodal Data - PMC, accessed May 16, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC11768470/
55. (PDF) Land Use Land Cover Classification with U-Net: Advantages of Combining Sentinel-1 and Sentinel-2 Imagery - ResearchGate, accessed May 16, 2025, https://www.researchgate.net/publication/354473786_Land_Use_Land_Cover_Classification_with_U-Net_Advantages_of_Combining_Sentinel-1_and_Sentinel-2_Imagery
56. tesi.luiss.it, accessed May 16, 2025, https://tesi.luiss.it/41874/1/770891_REZAEI_NILOOFAR.pdf
57. (PDF) Graph Neural Network for Daily Supply Chain Problems, accessed May 16, 2025,

https://www.researchgate.net/publication/384592034_Graph_Neural_Network_for_Daily_Supply_Chain_Problems
58. Graph Neural Network for Daily Supply Chain Problems - Preprints.org, accessed May 16, 2025, https://www.preprints.org/manuscript/202409.2376/v1
59. An Innovative Smart and Sustainable Low-Cost Irrigation System for Anomaly Detection Using Deep Learning - PubMed Central, accessed May 16, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC10892454/
60. (PDF) Detection of Anomaly using Machine Learning: A ..., accessed May 16, 2025, https://www.researchgate.net/publication/365060809_Detection_of_Anomaly_using_Machine_Learning_A_Comprehensive_Survey
61. (PDF) AI-Driven Risk Assessment Models for Financial Markets: Enhancing Predictive Accuracy and Fraud Detection - ResearchGate, accessed May 16, 2025, https://www.researchgate.net/publication/390162005_AI-Driven_Risk_Assessment_Models_for_Financial_Markets_Enhancing_Predictive_Accuracy_and_Fraud_Detection
62. Graph Neural Networks in Supply Chain Analytics and Optimization: Concepts, Perspectives, Dataset and Benchmarks - ResearchGate, accessed May 16, 2025, https://www.researchgate.net/publication/385786442_Graph_Neural_Networks_in_Supply_Chain_Analytics_and_Optimization_Concepts_Perspectives_Dataset_and_Benchmarks
63. [2411.08550] Graph Neural Networks in Supply Chain Analytics and Optimization: Concepts, Perspectives, Dataset and Benchmarks - arXiv, accessed May 16, 2025, https://arxiv.org/abs/2411.08550
64. accessed January 1, 1970, https.tracextech.com/minimize-eudr-risk-in-deforestation-free-supply-chain/
65. arxiv.org, accessed May 16, 2025, https://arxiv.org/pdf/2404.04305
66. AIRCentre/POS2IDON: Pipeline for ocean feature detection with Sentinel 2 - GitHub, accessed May 16, 2025, https://github.com/AIRCentre/POS2IDON
67. A Python Framework for Crop Yield Estimation Using Sentinel-2 ..., accessed May 16, 2025, https://www.mdpi.com/2673-4834/6/1/15
68. Sentinel-2 - Earth Engine Data Catalog - Google for Developers, accessed May 16, 2025, https://developers.google.com/earth-engine/datasets/catalog/sentinel-2
69. Sentinel-2 Cloud Masking with s2cloudless | Google Earth Engine, accessed May 16, 2025, https://developers.google.com/earth-engine/tutorials/community/sentinel-2-s2cloudless
70. samsammurphy/cloud-masking-sentinel2: Cloud masking of Sentinel 2 using Google Earth Engine - GitHub, accessed May 16, 2025, https://github.com/samsammurphy/cloud-masking-sentinel2
71. The Ultimate Guide to Tracking Deforestation Rates - Number Analytics, accessed May 16, 2025, https://www.numberanalytics.com/blog/ultimate-guide-tracking-deforestation-rates
72. REDD+ and climate finance, accessed May 16, 2025,

https://www.fao.org/forestry/our-focus/assessment-monitoring/en
73. Data Applications | World Resources Institute, accessed May 16, 2025, https://www.wri.org/data/data-applications
74. Conserving and Restoring Forests and Healthy Landscapes - World Resources Institute, accessed May 16, 2025, https://www.wri.org/forests
75. Remote sensing in forestry: current challenges, considerations and directions - Oxford Academic, accessed May 16, 2025, https://academic.oup.com/forestry/article/97/1/11/7159227
76. Navigating Data Challenges and Compliance for Deforestation-Free Supply Chains, accessed May 16, 2025, https://www.fao.org/in-action/forest-data-partnership/home/stories-detail/navigating-data-challenges-and-compliance-for-deforestation-free-supply-chains/en
77. Global Forest Watch - World Resources Institute, accessed May 16, 2025, https://www.wri.org/initiatives/global-forest-watch
78. Navigating traceability and the EUDR - Team Europe Initiative on Deforestation-free Value Chains, accessed May 16, 2025, https://zerodeforestationhub.eu/wp-content/uploads/2024/12/20241030_Traceability-Study-global-part_EUDR-Engagement_final.pdf
79. Top Illegal Deforestation Monitoring Companies - FlyPix AI, accessed May 16, 2025, https://flypix.ai/blog/illegal-deforestation-monitoring-companies/
80. Innovation Report in Deforestation-free Supply Chains - EMMA4EU, accessed May 16, 2025, https://www.emma4eu.eu/updates/innovation-report/
81. SEPAL Forest and Land Monitoring for Climate Action - Food and Agriculture Organization of the United Nations, accessed May 16, 2025, https://www.fao.org/in-action/sepal/certified-course/en
82. FAO's SEPAL homepage - Food and Agriculture Organization of the United Nations, accessed May 16, 2025, https://www.fao.org/in-action/sepal/en
83. (PDF) Global Forest Maps for the Year 2020 to Support the EU Regulation on Deforestation-free Supply Chains. Improved Map of Global Forest Cover (GFC2020) and Preliminary Map of Global Forest Types (GFT2020) - ResearchGate, accessed May 16, 2025, https://www.researchgate.net/publication/390583398_Global_Forest_Maps_for_the_Year_2020_to_Support_the_EU_Regulation_on_Deforestation-free_Supply_Chains_Improved_Map_of_Global_Forest_Cover_GFC2020_and_Preliminary_Map_of_Global_Forest_Types_GFT2020
84. Global Forest Watch: Forest Monitoring, Land Use & Deforestation Trends, accessed May 16, 2025, https://www.globalforestwatch.org/
85. accessed January 1, 1970, https://www.researchgate.net/publication/378603685_Deforestation_monitoring_in_the_Brazilian_Amazon_comparing_PRODES_DETER_and_other_global_datasets_Accuracy_and_implications_for_environmental_policy/fullTextFileContent
86. Coalition for Environmentally Sustainable Artificial Intelligence. | Élysée - elysee.fr, accessed May 16, 2025, https://www.elysee.fr/emmanuel-macron/2025/02/11/coalition-for-environmentally-sustainable-artificial-intelligence

87. Harnessing AI to Empower Smallholder Farmers: Bridging the Digital Divide for Sustainable Growth - Harvard ALI Social Impact Review, accessed May 16, 2025, https://www.sir.advancedleadership.harvard.edu/articles/harnessing-ai-empower-smallholder-farmers-bridging-digital-divide-sustainable-growth
88. Resolution-Agnostic Field Boundary Delineation on Satellite Imagery - arXiv, accessed May 16, 2025, https://arxiv.org/html/2504.02534v1
89. (PDF) AI4SmallFarms: A Dataset for Crop Field Delineation in ..., accessed May 16, 2025, https://www.researchgate.net/publication/374614935_AI4SmallFarms_A_Data_Set_for_Crop_Field_Delineation_in_Southeast_Asian_Smallholder_Farms
90. EUDR Compliance: 5 Ways Coffee Farms Use Geolocation Data - - Farmonaut, accessed May 16, 2025, https://farmonaut.com/remote-sensing/eudr-compliance-5-ways-coffee-farms-use-geolocation-data
91. (PDF) Early Identification of Crop Type for Smallholder Farming ..., accessed May 16, 2025, https://www.researchgate.net/publication/368313861_Early_Identification_of_Crop_Type_for_Smallholder_Farming_Systems_Using_Deep_Learning_on_Time-Series_Sentinel-2_Imagery
92. www.cifor-icraf.org, accessed May 16, 2025, https://www.cifor-icraf.org/publications/pdf_files/articles/AOkarda1801.pdf
93. Mapping of smallholder oil palm plantation and development of a growth model, accessed May 16, 2025, https://www.researchgate.net/publication/326727301_Mapping_of_smallholder_oil_palm_plantation_and_development_of_a_growth_model
94. Mapping the Distribution of Coffee Plantations from Multi-Resolution, Multi-Temporal, and Multi-Sensor Data Using a Random Forest Algorithm - MDPI, accessed May 16, 2025, https://www.mdpi.com/2072-4292/12/23/3933
95. Review of Remote Sensing Methods to Map Coffee Production Systems - MDPI, accessed May 16, 2025, https://www.mdpi.com/2072-4292/12/12/2041
96. Busting three myths about deforestation monitoring for the EUDR, accessed May 16, 2025, https://www.teaandcoffee.net/blog/36292/busting-three-myths-about-deforestation-monitoring-for-the-eudr/
97. Ina-Geoportal, accessed May 16, 2025, https://tanahair.indonesia.go.id/
98. Indonesia: One Map Policy - Open Government Partnership, accessed May 16, 2025, https://www.opengovpartnership.org/sites/default/files/case-study_Indonesia_One-Map-Policy.pdf
99. BRIN Provides Remote Sensing Satellite Imagery Data and ... - BRIN, accessed May 16, 2025, https://www.brin.go.id/en/news/117531/brin-provides-remote-sensing-satellite-imagery-data-and-information-services
100. Indonesian National Institute of Aeronautics and Space (LAPAN) - RPI - Landsat Missions, accessed May 16, 2025, https://landsat.usgs.gov/RPI

101. Fusion-Based Approaches and Machine Learning Algorithms for Forest Monitoring: A Systematic Review - MDPI, accessed May 16, 2025, https://www.mdpi.com/3042-4526/2/1/7
102. When Does Deep Learning Work Better Than SVMs or Random ..., accessed May 16, 2025, https://www.kdnuggets.com/2016/04/deep-learning-vs-svm-random-forest.html
103. DETER-R: An Operational Near-Real Time Tropical Forest Disturbance Warning System Based on Sentinel-1 Time Series Analysis - MDPI, accessed May 16, 2025, https://www.mdpi.com/2072-4292/14/15/3658
104. (PDF) SEMIAUTOMATIC AI-BASED REMOTE SENSING ..., accessed May 16, 2025, https://www.researchgate.net/publication/391572429_SEMIAUTOMATIC_AI-BASED_REMOTE_SENSING_CLASSIFICATION_TO_DELIVER_AN_EU_LANDMARK_2020_DEFORESTATION_MAP_FOR_BRAZIL
105. New Commodity Models to Support Deforestation-free Supply ..., accessed May 16, 2025, https://www.forestdatapartnership.org/news-events/new-commodity-models-to-support-deforestation-free-supply-chains
106. Statistical Comparison of Simple and Machine Learning Based Land Use and Land Cover Classification Algorithms: A Case Study - Journal of Water Management Modeling, accessed May 16, 2025, https://www.chijournal.org/H524
107. (PDF) Comparison of Machine and Deep Learning Models for the ..., accessed May 16, 2025, https://www.researchgate.net/publication/373528126_Comparison_of_Machine_and_Deep_Learning_Models_for_the_Prediction_of_Land_Degradation
108. Digitising land monitoring through GIS - Esri Indonesia, accessed May 16, 2025, https://esriindonesia.co.id/resources/news/digitising-land-monitoring-through-gis
109. accessed January 1, 1970, https://www.esri.com/en-us/industries/blog/sustainable-development/digitising-land-monitoring-through-gis/
110. Automating Land Cover Change Detection: A Deep Learning based approach to map deforested areas, accessed May 16, 2025, http://mtc-m21c.sid.inpe.br/col/sid.inpe.br/mtc-m21c/2020/06.09.11.59/doc/publicacao.pdf
111. Global Forest Monitoring from Earth Observation - OAPEN Library, accessed May 16, 2025, https://library.oapen.org/bitstream/20.500.12657/41670/1/9781466552029.pdf
112. Accuracy Assessment of the Global Forest Cover Map for the Year 2020: Assessment Protocol and Analysis - JRC Publications Repository, accessed May 16, 2025, https://publications.jrc.ec.europa.eu/repository/handle/JRC141231
113. (PDF) Deforestation and Forest Degradation in the Amazon -Update for year 2023 and assessment of humid forest regrowth - ResearchGate, accessed May 16, 2025, https://www.researchgate.net/publication/388218998_Deforestation_and_Forest_

Degradation_in_the_Amazon_-Update_for_year_2023_and_assessment_of_humid_forest_regrowth
114. Deforestation-Free Value Chains - FEFAC, accessed May 16, 2025, https://fefac.eu/wp-content/uploads/_pda/2024/07/2024-06-25-Portfolio-of-EUDR-workshop-presentations.pdf
115. How to Write a Successful Research Proposal? | Examples + ..., accessed May 16, 2025, https://proposally.ai/how-to-write-research-proposal/
116. Chapter 4. Preprocessing in remote sensing, accessed May 16, 2025, https://www.geo-informatie.nl/courses/grs10306/Clevers/RS%20CH4%20Preprocessing/RS%20Ch%204%20RS%20Preprocessing.pdf
117. Pre-processing of satellite images: atmospheric correction and cloud-masking - tethys.farm, accessed May 16, 2025, https://tethys.farm/en/atmospheric-cloud-masking-correction/
118. (PDF) A Web-Based Prototype System for Deforestation Detection on High-Resolution Remote Sensing Imagery With Deep Learning - ResearchGate, accessed May 16, 2025, https://www.researchgate.net/publication/382726117_A_Web-based_Prototype_System_for_Deforestation_Detection_on_High-resolution_Remote_Sensing_Imagery_with_Deep_Learning
119. AI-enabled forest monitoring: A new tool for combating deforestation - Picterra, accessed May 16, 2025, https://picterra.ch/blog/ai-enabled-forest-monitoring-a-new-tool-for-combating-deforestation/
120. 9 Accuracy Metrics to Evaluate AI Model Performance - Galileo AI, accessed May 16, 2025, https://www.galileo.ai/blog/accuracy-metrics-ai-evaluation
121. Field trip to Colombia. How Farmers Prepare for EUDR - Seedtrace, accessed May 16, 2025, https://seedtrace.org/blog/eudr-in-practice-with-farmers-at-the-origin/