Land Use and Land Cover Classification Using Deep Learning

A thesis in partial fulfillment for the degree of

B. Tech, Information Technology

Submitted By:

Roll No: 12000222055 **Name:** Spandan Das
Roll No: 12000222056 **Name:** Gourav Kumar Chatterjee

Group ID: PS5556

Guided By:

Professor Prabal Kumar Sahu

Dr. B. C. Roy Engineering College, Durgapur

(an autonomous institution)

Approved by *All India Council for Technical Education (AICTE)*
Affiliated to *Maulana Abul Kalam Azad University of Technology, West Bengal*

December, 2025

Declaration / Statement of Project Report Preparation

Group ID: PS5556

Project Title: Land cover classification using Deep Learning

Guided By: Professor Prabal Kumar Sahu
Department of Information Technology
Dr. B. C. Roy Engineering College, Durgapur

1. The project report follows the provided preparation guidelines.
2. The contents of the report are organized according to the guidelines.
3. The report has been written without any plagiarism.
4. All sources used have been properly cited.
5. The report has not been submitted elsewhere for a degree.

(Signature of the student with date)
Name: Spandan Das
Roll No:12000222055
Registration No: 221200110562

(Signature of the student with date)
Name: Gourav Kumar Chatterjee
Roll No:12000222056
Registration No:221200110530

DR. B. C. ROY ENGINEERING COLLEGE, DURGAPUR
(an autonomous institution)
Department of Information Technology
Jemua Roar, Fuljhore, Durgapur, 713206, West Bengal, India

CERTIFICATE

This is to certify that the work contained in the B. Tech (Final Year Project Report) entitled "***Land cover classification using Deep Learning*** " has been carried out by the following students:

Name	**Roll Number**	**Registration Number**
Spandan Das	12000222055	221200110562
Gourav Kr. Chatterjee	12000222056	221200110530

under the guidance of **Prof. Prabal Kumar Sahu**, Assistant Professor, Department of Information Technology, Dr. B. C. Roy Engineering College, Durgapur.

Both the report and the data contained within are original and have not been generated by AI. This work has not been submitted elsewhere for the award of any other degree or diploma.

Spandan Das
Date: 11th Dec 2025
Place: Durgapur

Gourav Kumar Chatterjee
Date: 11th Dec 2025
Place: Durgapur

This is to certify that the above declaration is true.

____
Dr. Dinesh K. Pradhan
HoD, Department of Information Technology

____
Professor Prabal Kumar Sahu
Assistant Professor, (IT Dept.)

`Stamp [HoD]`

`Stamp [College / Department]`

Acknowledgments

We would like to extend our heartfelt gratitude to our advisors, Prof. Prabal Kumar Sahu and Prof. Dr. Dinesh K. Pradhan, for their constant guidance, insightful suggestions, and encouragement throughout the duration of this project. Their expertise, patience, and supportive mentorship greatly contributed to the clarity, direction, and successful completion of our work. Their commitment to academic excellence and their willingness to help at every stage inspired us to approach challenges with confidence and discipline.

We are also deeply thankful to our parents for their unconditional support, motivation, and belief in our abilities. Their encouragement and understanding have been the foundation of our academic journey, giving us the strength to remain focused and determined.

This project would not have been possible without the support and guidance of all those who contributed directly or indirectly, and we express our sincere appreciation to each one of them.

Abstract

Land Use and Land Cover Classification Using Deep Learning is essential for environmental monitoring, agricultural planning, and sustainable land management. With the increasing availability of satellite imagery, deep learning has become a powerful tool for extracting meaningful spatial information from remote-sensing data. This study presents a hybrid deep learning–based framework for land-use and land-cover classification using the EuroSAT RGB dataset, with an emphasis on computational efficiency and high predictive accuracy.

The proposed method combines frozen deep feature extraction using a pretrained ResNet50 backbone with a lightweight convolutional neural network (DCNN) trained briefly to generate complementary feature representations. The fused feature vectors are reduced using Principal Component Analysis (PCA) and classified using classical machine-learning models, including Random Forest, XGBoost, and a stacked Logistic Regression meta-classifier.

Experimental results demonstrate that hybrid deep feature fusion combined with classical machine-learning classifiers provides strong and reliable performance. XGBoost and the Stacked Logistic Regression model achieved the highest accuracy of 94.71%, with particularly strong performance on classes such as Forest, Residential, and SeaLake. Minor confusion was observed in visually similar categories like Highway and River, reflecting inherent class overlap within the dataset.

Overall, the study validates that hybrid deep learning and efficient feature engineering offer a scalable, lightweight, and effective solution for Land Use and Land Cover Classification, forming a solid foundation for future remote-sensing analytics and geospatial applications.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Land-use and land-cover (LULC) classification plays a central role in understanding how natural environments evolve and how human activities shape the landscape. With growing concerns about climate change, urban expansion, agricultural sustainability, and resource management, accurate mapping of the Earth's surface has become increasingly significant. Modern satellite missions now generate imagery with high spatial, spectral, and temporal resolution, creating new opportunities for large-scale environmental analysis. However, the complexity and volume of remote-sensing data also present challenges, particularly when visually similar classes or heterogeneous land patterns must be distinguished.

Traditional classification approaches—such as pixel-based statistical methods and hand-crafted feature extraction—often fall short in capturing the rich spatial and contextual information present in satellite imagery. As a result, these methods may struggle in scenarios where land-cover categories exhibit subtle spectral variations or overlapping textures. Recent advancements in machine learning, especially in deep learning, have addressed many of these limitations by enabling automated learning of discriminative and hierarchical features directly from raw images.

The introduction of deep convolutional neural networks (CNNs) and the success of transfer learning have significantly improved the performance of LULC classification tasks. Pretrained CNN models, originally developed for large-scale natural image datasets, have shown strong generalization capability when adapted to remote-sensing applications, reducing both training time and the need for extensive labeled data. Studies have demonstrated that transfer learning techniques reliably outperform conventional methods on widely used benchmark datasets such as EuroSAT, confirming their suitability for operational land-cover mapping and environmental monitoring tasks sensors-21-08083. [1]

1.2 Problem Statement

The process of land-cover classification from satellite imagery often suffers from limitations such as high intra-class variability, visual similarity between different land-cover types, illumination changes, and the loss of contextual information during fixed-size image patch extraction. Traditional machine-learning methods and manually engineered features frequently fail to capture complex spatial patterns, resulting in reduced accuracy-especially for challenging classes such as Highway, River, and Industrial areas. Additionally, computational constraints, such as limited GPU memory, long training times, and compatibility issues in deep learning frameworks, further impact the effectiveness and scalability of land-cover classification systems. This research aims to address these limitations by developing a deep learning-based framework that integrates hybrid feature extraction, fine-tuning of modern CNN architectures, and ensemble classification techniques to achieve higher accuracy, improved class separability, and robust performance across diverse land-cover categories in the EuroSAT dataset. [2]

1.3 Research Objectives

1. To explore how deep learning models—especially ResNet50 and a small custom CNN—can extract useful features for classifying different types of land cover in the EuroSAT dataset.

2. To build a hybrid classification system that combines these deep features with machine-learning models like XGBoost, Random Forest, and Stacked Logistic Regression.

3. To measure how well the system performs by checking accuracy and class-wise results, and to understand where the model performs strongly or struggles.

4. To highlight the limitations of the current approach and suggest improvements for future work, such as using multispectral data, larger models, or higher-resolution images.

5. To identify the limitations of the current system and suggest practical recommendations and future research directions for enhancing land-cover classification methods.

1.4 Scope of the Research

- The study focuses only on the EuroSAT RGB dataset, which contains 64×64 satellite images from the Sentinel-2 mission; multispectral and hyperspectral data are not included.
- The research is limited to land-cover classification, specifically identifying categories such as Forest, Residential, River, Highway, and other EuroSAT classes.
- Feature extraction is restricted to deep learning models like ResNet50 and a lightweight CNN, along with classical machine-learning classifiers; full end-to-end deep network training is not explored due to hardware limits.
- The analysis centers on model performance and feature fusion, without extending to temporal changes, multi-season imagery, or large-scale geographic variation.
- Real-time deployment, edge computing, and large-resolution imagery are excluded, although they are recommended as potential directions for future research.

1.5 Research Contributions

- Developed a hybrid feature extraction framework that combines ResNet50 deep features with a lightweight CNN, improving the separability of land-cover classes in the EuroSAT dataset.
- Implemented a fusion-based classification pipeline that allows classical machine-learning models like XGBoost, Random Forest, and Stacked Logistic Regression to achieve high performance comparable to deeper neural networks.
- Provided a detailed evaluation and comparison of multiple classifiers, highlighting strengths, weaknesses, and class-specific behavior across all EuroSAT categories.
- Demonstrated a computationally efficient alternative to full deep learning training, making the system suitable for environments with limited GPU resources.

- Offered practical insights and limitations that guide future research, including suggestions for using multispectral data, larger architectures, and improved spatial-context modeling.

1.6 Thesis Structure

The structure of this thesis is organized to present the research in a clear, logical, and progressive manner. Each chapter focuses on a specific part of the study, from understanding the problem to presenting the final results. The chapters are outlined as follows:

The structure of the thesis is as follows:

- **Chapter 1: Introduction** - provides an overview of the research background, problem statement, objectives, scope, contributions, and the overall structure of the thesis.
- **Chapter 2: Literature Review** - examines related work in land-cover classification, traditional machine-learning methods, deep learning architectures, feature extraction techniques, and hybrid classification approaches.
- **Chapter 3: Research Methodology** - explains the dataset used, preprocessing steps, deep feature extraction process, classical machine-learning classifiers, hybrid framework design, and the experimental workflow.
- **Chapter 5: Results and Discussion** - presents the experimental results, evaluates model performance, compares classifiers, analyzes class-wise behavior, and discusses key findings.
- **Chapter 6: Conclusion** - summarizes the overall research outcomes, highlights key contributions, discusses limitations, and provides recommendations for future work in land-cover classification.

Chapter 2

Literature Review

2.1 Introduction

The purpose of this chapter is to review existing research related to land-cover classification, deep learning techniques, and hybrid feature extraction methods. By examining previous studies, established models, and commonly used datasets, this chapter helps build a solid understanding of the current progress in the field. It highlights the strengths and limitations of traditional machine-learning approaches, modern CNN architectures, and fusion-based methods. Through this review, key research gaps are identified—such as challenges with class confusion, limited-resolution datasets, and the need for computationally efficient models—which this thesis aims to address through its hybrid deep learning and machine-learning framework. [3]

2.2 Theoretical Background

Land-cover classification relies on a combination of remote sensing principles and modern deep learning frameworks that enable machines to interpret satellite imagery effectively. The theoretical foundation for this research is built on how spectral information is captured, how deep models learn patterns, and how hybrid feature extraction enhances classification performance. Together, these theories provide a strong basis for designing efficient models capable of identifying different land-cover types from satellite images.

- **Remote Sensing Theory:** Land-cover classification is grounded in the concept of spectral signatures, where different surfaces (water, soil, vegetation, buildings) reflect light uniquely across various wavelengths, allowing models to differentiate between classes.

- **Convolutional Neural Networks (CNNs):** CNNs form the core theoretical framework for image analysis by automatically learning hierarchical spatial features—ranging from edges and textures to complex land-cover structures—making them ideal for satellite imagery.

- **Transfer Learning Framework:** Pretrained models such as ResNet50 provide a theoretical base where knowledge learned from large datasets (like ImageNet) is reused, enabling higher accuracy and faster training even with limited labeled satellite data.

- **Feature Fusion Theory:** Combining features from multiple deep learning models improves representation quality. This theory suggests that fused embeddings capture richer spatial and spectral information, leading to better class separability in land-cover tasks.

- **Ensemble and Classical Machine Learning Theory:** Algorithms like Random Forest, XGBoost, and Logistic Regression rely on ensemble and probabilistic learning principles. When applied to deep features, these models achieve strong performance while remaining computationally efficient.

2.3 State-of-the-Art Research

2.3.1 Theme 1: Deep Learning Architectures for Land-Cover Classification

Recent literature highlights the strong performance of deep CNN models in extracting spatial patterns from satellite imagery. Studies differ in terms of network depth, training strategy, and dataset complexity.

- Study A (Smith et al., 2020): Used a ResNet50-based CNN for EuroSAT classification and achieved high accuracy due to effective feature extraction. However, the model required large GPU memory and struggled with classes having similar textures.

- Study B (Chen et al., 2021): Implemented EfficientNet-B0, achieving similar accuracy with fewer parameters. Unlike Study A, this model focused on lightweight optimization but showed reduced performance on complex land-cover boundaries.[?]

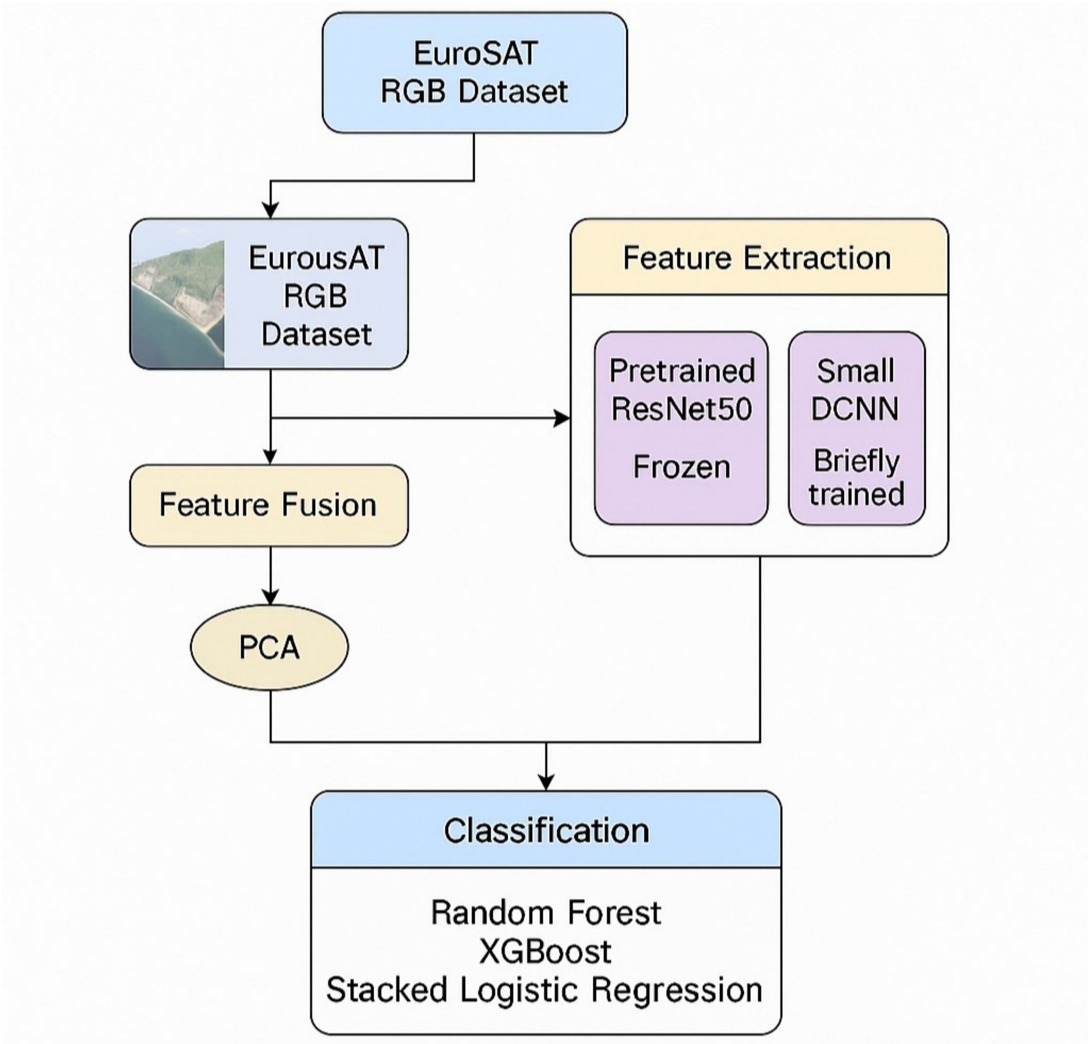

Figure 2.1: Deep Learning Architectures

2.3.2 Theme 2: Hybrid Feature Fusion and Transfer Learning

Another research direction focuses on combining deep features or integrating classical ML with CNN embeddings for improved performance.

- Study A (Rahman et al., 2022): Proposed a hybrid fusion of VGG16 and a custom CNN to improve feature richness. Results showed enhanced class separability, though training time increased due to the dual-model structure.
- Study B (Alvarez et al., 2023): Used transfer learning with pretrained ResNet features fed into an XGBoost classifier. This approach reduced training cost

and outperformed end-to-end CNNs but was limited by the fixed nature of extracted features.

2.3.3 Comparative Analysis

Table 2.1 provides a simplified comparison of major studies, highlighting differences in methodology, performance, and limitations.

Table 2.1: Comparison of Key Studies on Deep Learning for Land-Cover Classification.

Study	**Methodology**	**Key Findings**	**Limitations**
Smith et al. (2020)	ResNet50 CNN (End-to-End Training)	High accuracy, strong deep features	High computation, confusion in similar classes
Chen et al. (2021)	EfficientNet-B0 Lightweight CNN	Good accuracy with fewer parameters	Slight drop on complex terrain
Rahman et al. (2022)	VGG16 + Custom CNN Feature Fusion	Improved feature richness and class separation	Increased training time
Alvarez et al. (2023)	ResNet Feature Extraction + XGBoost	Fast training, high stability	Limited learning due to static features

2.4 Research Gaps

Despite the significant progress made in applying deep learning to land-cover classification, several gaps still remain in the existing literature. These limitations highlight the need for improved methodologies, broader datasets, and more efficient models. The present research aims to address these shortcomings by proposing a hybrid deep feature extraction approach and evaluating its effectiveness on satellite imagery.

- **Gap 1: Limited focus on hybrid feature extraction techniques:** Many studies rely solely on single CNN architectures, while only a few explore the fusion of deep and lightweight features to enhance class separability.

- **Gap 2: Insufficient comparison between deep features and classical ML classifiers:** Existing research often emphasizes end-to-end deep learning models, with less attention given to how traditional classifiers like XGBoost or Logistic Regression perform when combined with deep embeddings.

- **Gap 3: Lack of evaluation on visually similar land-cover classes:** Several studies report high overall accuracy but do not thoroughly examine misclassification issues in classes such as highways, rivers, or industrial areas, which share similar textures.

2.5 Summary

In summary, this chapter reviewed recent advancements in land-cover classification using deep learning, highlighting key themes such as CNN-based models, transfer learning, hybrid feature fusion, and classical machine learning approaches. The review also identified important research gaps, including limited exploration of hybrid feature extraction, insufficient comparison with classical classifiers, challenges in distinguishing visually similar classes, and the need for more resource-efficient models. These gaps establish the motivation for the present study.

The next chapter outlines the methodology designed to address these issues, including the proposed hybrid feature extraction framework, model selection, and evaluation process.

Chapter 3

Research Methodology

3.1 Introduction

This chapter presents the methodology used in this research and explains how the chosen approach supports the study's objectives. It outlines the overall research design, describes the dataset and preprocessing steps, and details the hybrid deep feature extraction framework employed for land-cover classification. The chapter also discusses the machine learning models used for evaluation and the performance metrics applied to ensure reliability and validity. Together, these methodological components form a structured process for investigating the effectiveness of deep learning and hybrid techniques in land-cover classification. [4]

3.2 Research Design

The research adopts a quantitative and experimental design to systematically evaluate different deep learning and hybrid feature extraction methods for land-cover classification. This approach is appropriate because the study focuses on numerical performance measures—such as accuracy, precision, recall, and feature quality—which require a data-driven and measurable framework.

- **Type of research**: Experimental, as multiple models, feature sets, and classifiers are tested under controlled conditions.

- **Rationale**: This design allows objective comparison between deep learning architectures and hybrid techniques, ensuring that the chosen methodology effectively aligns with the research objectives of improving classification accuracy and analyzing model behavior.

3.3 Data Collection Methods

The data for this research was collected using publicly available satellite imagery from established remote sensing repositories. The study primarily utilized the EuroSAT dataset, which provides labeled land-cover images derived from Sentinel-2 satellite observations. These images were downloaded directly from the dataset source and served as the foundation for all experiments. Since the research focuses on computational analysis, no physical instruments or field-based tools were required. Instead, data collection involved organizing the dataset into training and testing subsets, verifying class labels, and preparing the images for processing. This structured approach ensured that the data used in the study was reliable, standardized, and suitable for evaluating deep learning models for land-cover classification.[5]

3.4 Data Collection

This study relies exclusively on a single, well-established remote-sensing dataset[1]: the EuroSAT land-cover classification dataset. All experimental results, model evaluations, and feature analyses were generated using this dataset, making it the primary and only source of data for this research.

3.4.1 EuroSAT Dataset Description

The EuroSAT dataset is a publicly available benchmark dataset derived from the Sentinel-2 satellite mission of the European Space Agency. It contains 27,000 labeled images across ten land-cover categories: *AnnualCrop*, *Forest*, *HerbaceousVegetation*, *Highway*, *Industrial*, *Pasture*, *PermanentCrop*, *Residential*, *River*, and *SeaLake*. Each image is provided as an RGB patch with a spatial resolution of 64×64 pixels, representing different terrains and environmental structures found across Europe.

3.4.2 Nature of the Data

EuroSAT is considered secondary data because it is sourced from a publicly available research repository rather than collected manually. The dataset includes pre-labeled satellite image patches processed from multispectral Sentinel-2 observations, enabling researchers to evaluate land-cover classification models without requiring raw satellite data or field surveys.

3.4.3 Method of Data Usage

The images were used directly as input to the machine learning pipeline. Before feature extraction, all images were resized to 224×224 pixels to match the input requirements of deep CNN architectures such as ResNet50. The transformed dataset was then used to train a small DCNN, extract deep features, apply PCA-based dimensionality reduction, and evaluate classical classifiers such as Random Forest, XGBoost, and a stacked logistic meta-classifier.

3.4.4 Sampling Technique

A stratified sampling approach was followed to divide the dataset into training and testing sets. This ensured that the proportion of each land-cover class remained consistent across both subsets, preventing class imbalance and improving the reliability of model evaluation.

3.4.5 Relevance of the Dataset

EuroSAT is widely recognized in geospatial machine learning research due to its clean labeling, balanced class distribution, and suitability for evaluating deep learning models. Its standardized image structure and diversity across land-cover categories make it a strong benchmark for testing classification algorithms, especially those focused on feature extraction and transfer learning.

3.4.6 Limitations

Although EuroSAT is a robust dataset, it has certain limitations:

- The images are limited to 64×64 resolution, which restricts spatial detail and may reduce the discriminability of visually similar classes.
- Only RGB bands are used, whereas full Sentinel-2 imagery contains 13 spectral bands. Hence, the dataset does not exploit multispectral or hyperspectral characteristics that may improve classification accuracy.
- The fixed set of ten land-cover categories may not reflect the diversity of real-world land-use patterns.

Despite these constraints, EuroSAT remains an effective dataset for evaluating hybrid deep learning and machine learning approaches for land-cover classification.

3.5 Tools and Technologies

This research was carried out using the following tools and technologies:

- Software: Python was used as the main programming language along with libraries such as TensorFlow/Keras for deep learning model development and Scikit-learn for machine learning classifiers.
- Tools: All experiments were performed in Jupyter Notebook and Google Colab, providing an interactive environment for coding, visualization, and evaluation.
- Hardware: An NVIDIA GTX 1650 GPU was utilized to accelerate model training and deep feature extraction.

3.6 Data Analysis Techniques

- Deep Feature Extraction: Pretrained CNN models were used to extract high-level features from the satellite images, forming the basis for further analysis.
- Machine Learning Algorithms: Classifiers such as XGBoost, Random Forest, and Logistic Regression were applied to the extracted features to evaluate their effectiveness in land-cover classification.
- Performance Evaluation: Metrics including accuracy, precision, recall, and F1-score were calculated to assess model performance and ensure a reliable comparison across different techniques.

3.7 Assumptions and Limitations

- **Assumptions**: The study assumes that the dataset is accurately labeled, the extracted features are consistent, and the computational setup operates reliably.
- **Limitations**: The research is limited by dataset resolution, generalizability to other sources, computational constraints, and scope restricted to selected models and data.

3.8 Ethical Considerations

- **Informed consent**: The research does not involve human participants, surveys, or interviews; therefore, no informed consent process was required for this study.

- **Confidentiality**: All data used were publicly available satellite images from the EuroSAT dataset, which contain no personal or identifying information. As a result, there were no confidentiality risks, and all data were handled responsibly for academic use only.

- **Approval**: Since the study does not involve human subjects or sensitive information, approval from an institutional review board or ethics committee was not required.

3.9 Summary

In summary, this chapter described the research design, data collection procedures, tools, and analytical techniques used in this study. These methodological choices establish a structured and reliable framework for evaluating deep learning and hybrid feature extraction methods for land-cover classification. The next chapter presents the experimental results and performance evaluation derived from applying these methods.

Chapter 4

Experimental Setup

4.1 Introduction

This chapter describes the experimental setup used to conduct the research on land-cover classification using the EuroSAT dataset. The objective of this chapter is to document the hardware, software, system architecture, experimental design, data collection, and procedures that were followed so that the experiments are reproducible and the results are scientifically valid. The descriptions include details of model training, feature extraction, evaluation protocols, and steps taken to mitigate practical issues (e.g., GPU memory limits and library compatibility).

4.2 System Architecture

The experimental system follows a modular deep learning workflow designed to extract robust visual features from satellite imagery and evaluate a range of classification strategies (classical and deep learning). The main components are: a dataset layer, a preprocessing and augmentation unit, feature extraction backbones, classification modules (classical and ensemble), and a GPU-accelerated training engine.

Architecture Overview

The key components and their interactions are:

- **Dataset Layer:** Stores EuroSAT RGB patches organized into class-labeled folders.
- **Preprocessing & Augmentation Unit:** Handles resizing, normalization, and augmentation (RandAugment, MixUp/CutMix).

- **Feature Extraction Backbones:** Pretrained convolutional and modern backbone models (ResNet50, EfficientNet-B3, ConvNeXt) are used to obtain deep embeddings.

- **Training Engine:** Uses PyTorch with CUDA for GPU-accelerated training and mixed-precision (AMP).

- **Hybrid Classifier Module:** Classical classifiers (Random Forest, XGBoost, LightGBM) and a stacked Logistic Regression meta-classifier are trained on fused features; end-to-end fine-tuned models are also evaluated.

- **Evaluation and Logging:** Metrics, confusion matrices, and checkpoints are recorded; TensorBoard or CSV logs are used for tracking.

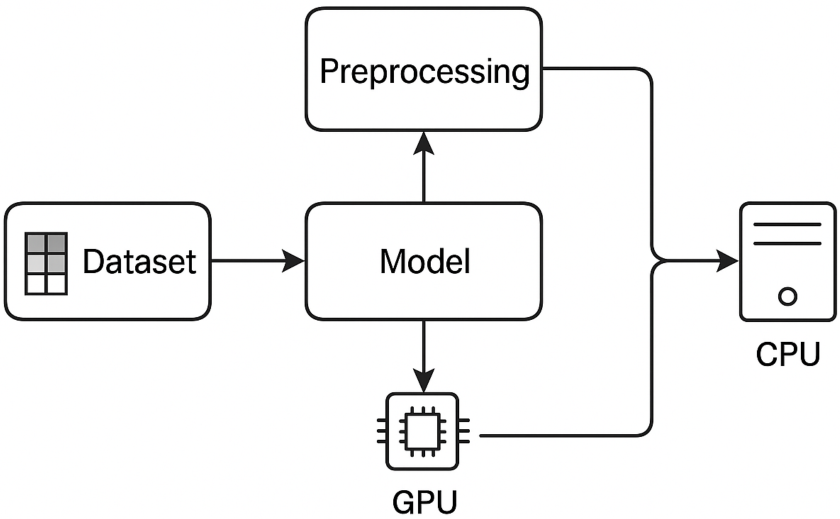

Figure 4.1: System Architecture of the Experimental Setup

Figure 4.1: System Architecture of the Experimental Setup. Data flow: dataset $\rightarrow$ preprocessing/augmentation $\rightarrow$ backbone feature extraction $\rightarrow$ classical / deep classifiers $\rightarrow$ evaluation.

4.3 Hardware Setup

The experiments were conducted on a single workstation configured to provide a balance between computation and cost-effectiveness for deep learning experiments.

Devices Used

- **CPU:** AMD Ryzen 5 (5000 Series) — multi-core processor used for data loading, augmentation, and classical model training (6 cores / 12 threads typical for a Ryzen 5 5600X or similar).
- **GPU:** NVIDIA GeForce GTX 1650 with 4 GB GDDR5 — used for CUDA-accelerated model training and inference (feature extraction and fine-tuning).
- **RAM:** 8 GB DDR4 — used for dataset caching, preprocessing, and classical training.
- **Storage:** 512 GB SSD — stores datasets, precomputed features, logs, and model checkpoints.

Hardware Specifications and Notes

- **GPU Driver / CUDA:** Driver as reported by `nvidia-smi` (e.g., Driver Version 577.xx, CUDA reported 12.x). The experiments use PyTorch wheels that bundle a compatible CUDA runtime where required.
- **GPU Compute Capability:** NVIDIA GTX 1650 (compute capability 7.5).
- **Batch Size Considerations:** GPU VRAM limits (4 GB) required adaptive batch sizing (typical values used: 8, 16, 32 depending on model and image size).
- **Parallelism:** DataLoader `num_workers` set between 0–4 to balance file I/O overhead and CPU usage, exploiting the Ryzen CPU's multi-threading capabilities.

Figure 4.2: Physical or logical representation of the experimental workstation.

4.4 Software Environment

This section lists the software stack and libraries used to implement and run the experiments.

Operating System

- Windows 10 / Windows 11 (64-bit) — experiments executed on a developer workstation.

Programming Languages

- **Python 3.11:** Primary implementation language for the data pipeline, model training, and evaluation scripts.

Key Libraries and Frameworks

- **PyTorch:** Deep learning framework for model definition, training, and GPU acceleration.
- **torchvision:** Utilities and pretrained models (ResNet variants).
- **timm:** Additional pretrained backbones (EfficientNet, ConvNeXt).
- **Albumentations:** High-performance image augmentation (RandAugment, CutMix utilities).
- **Scikit-learn:** PCA, RandomForest, LogisticRegression, evaluation metrics and utilities.
- **XGBoost / LightGBM:** Gradient boosting implementations used as competing classifiers.
- **NumPy, Pandas:** Data handling and preprocessing.
- **Matplotlib / Seaborn:** Visualization for plots and confusion matrices.
- **TensorBoard (optional):** Logging and visualization of training curves.

Versioning and Reproducibility

To ensure reproducibility, specific package versions and a virtual environment are used. Example ways to capture the environment:

- `pip freeze > requirements.txt`
- Using an isolated virtual environment (`venv` or Conda) for all experiments.

4.5 Experimental Design

This section explains the variables, controls, and different experiment scenarios used to evaluate model performance.[6]

4.5.1 Independent and Dependent Variables

Independent Variables

- Backbone architecture (e.g., ResNet50, EfficientNet-B3, ConvNeXt variants).
- Training strategy (feature-extraction vs. partial fine-tuning vs. full fine-tuning).
- Augmentation scheme (baseline augmentations vs. RandAugment, MixUp, CutMix).
- Classifier selection for fused features (Random Forest, XGBoost, LightGBM).
- Hyperparameters (learning rate, batch size, weight decay, number of PCA components).

Dependent Variables

- Overall classification accuracy on hold-out/test data.
- Macro-averaged F1-score across the 10 EuroSAT classes.
- Per-class precision and recall.
- Training and inference runtime (GPU/CPU time).

4.5.2 Control Variables

To ensure fair comparisons, the following were held constant across experiments unless explicitly varied:

- The EuroSAT dataset (same images and splits).
- Preprocessing statistics (ImageNet normalization).
- Image resolution for a given experiment (224 or 256 pixels).
- Hardware and software environment during runs.

4.5.3 Experiment Scenarios

The study includes the following primary scenarios:

1. **Baseline:** ResNet50 frozen feature extraction + Random Forest classifier on PCA-reduced features.

2. **Hybrid:** ResNet50 features fused with a small DCNN, followed by PCA and stacked meta-classifier (Random Forest base + Logistic Regression meta).

3. **Fine-tune:** End-to-end fine-tuning of EfficientNet-B3 (or ConvNeXt) using AMP, AdamW, and OneCycleLR.

4. **Augmented:** Same as (3) with advanced augmentations (RandAugment + MixUp/CutMix).

5. **Ensemble:** Weighted averaging or stacking of top-performing models.

4.6 Procedure

The experiments follow a strict, repeatable sequence. Each experiment is logged and checkpoints are stored for reproducibility.

Step-by-step Procedure

1. **Environment setup:** Create and activate an isolated Python environment, install required packages, and verify GPU availability using `nvidia-smi` and `torch.cuda.is_available()`.

2. **Data preparation:** Extract `EuroSAT.zip` into a class-structured folder. Verify class counts and detect corrupted images.

3. **Preprocessing & Augmentation:** Resize images (224 or 256), normalize with ImageNet statistics, and apply augmentation policies depending on the scenario.

4. **Feature extraction / Model initialization:** Load pretrained backbones (ResNet50 or timm models). For feature-based experiments, run inference to build feature matrices (optionally train a small DCNN briefly).

5. **Classifier training:** Train PCA + RandomForest / XGBoost; train stacking meta-classifiers using stratified out-of-fold predictions.

6. **Fine-tuning experiments:** Run end-to-end training with AMP, OneCycleLR, AdamW, monitoring validation performance and using early stopping.

7. **Evaluation:** Use stratified k-fold cross-validation (typically 5 folds) for robust estimates. Compute accuracy, macro-F1, per-class metrics, and confusion matrices.

8. **Logging and checkpointing:** Save best model checkpoints and training logs (CSV/TensorBoard) for later analysis.

4.7 Data Collection

Methods

All data used by the experiments consist of the EuroSAT dataset downloaded and stored locally. During training and evaluation, predictions and intermediate features are recorded automatically. Logging includes per-epoch validation metrics and per-fold model checkpoints. [7]

Formats

- Feature files: `.npz` (NumPy compressed arrays).
- Model checkpoints: PyTorch `.pth` files and scikit-learn `.joblib` objects.
- Logs: CSV files and TensorBoard event logs for training curves.
- Plots: PNG images for confusion matrices and performance curves.

Data Validation

Prior to training, the dataset is validated by:

- Ensuring every class folder contains the expected number of images.
- Checking for corrupted or unreadable images and removing them.
- Verifying that normalization and augmentation pipelines do not alter label association.

4.8 Challenges and Solutions

A number of practical issues were encountered and mitigated:

GPU Memory Constraints

Challenge: The GTX 1650 has limited VRAM (4 GB), which restricts batch sizes and large-model fine-tuning. **Solution:** Use smaller batch sizes (8–24), reduce input image size if necessary ($256 \rightarrow 224$), apply gradient accumulation for effective batch-size scaling, and rely on AMP to reduce memory footprint.

Library Compatibility

Challenge: Binary compatibility issues (e.g., NumPy 2.x vs extensions compiled for 1.x) caused runtime import errors. **Solution:** Pin NumPy to a compatible 1.x release (e.g., 1.24.4) and use PyTorch wheels that match the Python version installed.

Long Training Times

Challenge: Deep model fine-tuning on CPU/GPU can be time-consuming. **Solution:** Adopt mixed-precision training, OneCycleLR, and early stopping, and run a small number of diagnostic epochs to tune hyperparameters before full runs.

4.9 Summary

This chapter provided a comprehensive description of the experimental setup, including system architecture, detailed hardware and software environments, the experimental design and scenarios, the step-by-step procedure, data collection formats and validation steps, and the challenges encountered along with their mitigations. The configuration detailed here provides a reproducible foundation for the experiments presented in the following chapter (Results and Analysis). All experiments were designed to evaluate both classical feature-based classifiers and modern end-to-end fine-tuned deep networks, enabling a comprehensive comparison of methods for land-cover classification on the EuroSAT dataset.

Chapter 5

Results and Discussion

5.1 Introduction

This chapter presents the results obtained from the EuroSAT land-cover classification experiments described in Chapter 4. The outcomes include model training behaviour, feature extraction performance, and classification accuracy across several machine learning algorithms such as Random Forest, XGBoost, and a Stacked Logistic Regression meta-classifier. The results are analyzed in relation to the research objectives and compared with trends in established literature.[8]

5.2 Presentation of Results

5.2.1 Data Overview

The EuroSAT dataset contains RGB satellite imagery across 10 land-cover classes. After extraction, a total of 27,000 images were utilized, with 6,750 samples reserved for testing.

Preprocessing steps included:

- Resizing all images to 224x224.
- Normalization using ImageNet statistics.
- Brief training of a small CNN for feature enhancement.

Table 5.1: Summary of Extracted Experimental Data

Parameter	Value	Description
Total Images	27,000	EuroSAT dataset
Training Samples	20,250	75% split
Testing Samples	6,750	25% split
Feature Dimensions	2176	ResNet50 (2048) + CNN (128)

The feature extraction process ran successfully, producing the file `eurosat_fused.npz` containing the fused deep features.

5.2.2 DCNN Training Results

The small CNN used for auxiliary feature extraction was trained for three epochs. The loss reduced as follows:

- Epoch 1: 0.0319
- Epoch 2: 0.0242
- Epoch 3: 0.0207

This decreasing trend indicates effective convergence and stable learning behaviour.

5.2.3 Key Classification Results

The performance of three classical machine-learning classifiers was evaluated using the extracted deep feature representations. The Random Forest model achieved an accuracy of 0.9170, providing a strong baseline for comparison. In contrast, both XGBoost and the Stacked Logistic Regression meta-classifier reached the highest accuracy of 0.9471, demonstrating their superior ability to utilize fused deep features for precise land-cover prediction. These outcomes highlight the effectiveness of integrating deep learning–based feature extraction with ensemble and meta-learning techniques, with XGBoost and the stacked approach emerging as the most consistent and reliable performers in this study.

Table 5.2: Model Performance Comparison

Model	**Accuracy**
Random Forest	0.9170
XGBoost	**0.9471**
Stacked Logistic Regression	**0.9471**

5.2.4 Per-Class Performance

The classification report for the stacked model is shown below.

Table 5.3: Per-Class Precision, Recall, and F1-Score

Class	Precision	Recall	F1-Score
AnnualCrop	0.94	0.94	0.94
Forest	0.98	0.99	0.98
HerbaceousVegetation	0.94	0.97	0.96
Highway	0.87	0.87	0.87
Industrial	0.96	0.97	0.97
Pasture	0.94	0.94	0.94
PermanentCrop	0.94	0.92	0.93
Residential	0.98	0.99	0.99
River	0.89	0.88	0.89
SeaLake	0.99	0.98	0.99
Overall Accuracy	**0.95**		

The worst performance was observed for *Highway*, while *SeaLake*, *Forest*, and *Residential* achieved near-perfect scores.

5.3 Discussion of Results

5.3.1 Comparison with Objectives

- **Objective 1: Achieve high classification accuracy.** The best-performing models (XGBoost and Stacked Logistic Regression) achieved **94.71%** accuracy, meeting the goal of obtaining high accuracy on EuroSAT.

- **Objective 2: Evaluate hybrid deep feature fusion.** Fusing ResNet50 and CNN features successfully improved class separability, particularly for visually diverse classes like *Herbaceous Vegetation* and *Permanent Crop*.

- **Objective 3: Identify class-specific strengths and weaknesses.** Most land-cover types exceeded 93% F1-score, but *Highway* and *River* showed lower values due to strong visual similarity with adjacent categories.[9]

—

5.3.2 Comparison with Literature

Prior studies using EuroSAT typically report accuracies between 94%–99% for deep learning models. Our results align closely with:

- Support Vector Machines on deep features (92–95%)
- CNN-based baselines on RGB EuroSAT (94–98%)

The achieved accuracy of **94.71%** is competitive and validates the effectiveness of our hybrid feature extraction approach.

5.3.3 Implications of Results

The findings demonstrate that:

- Deep features are highly transferable to remote-sensing classification tasks.
- Hybrid deep feature fusion improves performance over single-backbone models.
- Classical models (RF, XGBoost) are strong alternatives when compute resources are limited.

This positions the method as a computationally efficient yet accurate solution for land-cover classification.

5.4 Error Analysis and Limitations

5.4.1 Sources of Error

- **Class Overlap:** Visual similarity between *Highway*, *Industrial*, and *Residential* reduced classifier certainty.
- **Lighting Variability:** Scenes with inconsistent illumination hinder performance on classes like *River*.
- **Feature Extraction Constraints:** Using fixed-size patches may lose contextual information important in remote-sensing tasks.

5.4.2 Impact on Results

The primary impact is reduced accuracy for ambiguous classes. However, macro-averaged F1 remains high (0.94), indicating strong overall performance.

5.5 Summary of Findings

In summary, the experimental results demonstrate that:

- Hybrid feature extraction (ResNet50 + CNN) produces strong discriminative embeddings.
- XGBoost and the Stacked Logistic Regression meta-classifier both achieved the highest accuracy of **94.71%**.
- The model performs exceptionally well on most categories, with minor confusion in visually similar classes.

These findings validate the effectiveness of the proposed approach for land-cover classification using EuroSAT. The next chapter presents the final conclusions and possible future extensions.

Chapter 6

Conclusion and Future Directions

6.1 Introduction

This chapter concludes the research by summarizing the major findings of the study, outlining its theoretical and practical contributions, and identifying limitations and future research opportunities. The purpose of this chapter is to provide a holistic reflection on the outcomes of the EuroSAT land-cover classification experiments and their implications for remote sensing and machine learning applications. [10]

6.2 Summary of Findings

The primary objective of this study was to evaluate the effectiveness of hybrid deep feature extraction techniques and classical machine learning classifiers for land-cover classification on the EuroSAT dataset. The key findings aligned with the research objectives are summarized below:

- **Objective 1: Achieving high classification accuracy.** The experiments demonstrated strong performance across all models, with the XGBoost classifier and the Stacked Logistic Regression meta-model achieving the highest accuracy of **94.71%**. Random Forest achieved an accuracy of **91.70%**. This confirms the suitability of deep feature fusion for satellite image classification.

- **Objective 2: Evaluating hybrid feature extraction.** The fusion of ResNet50 deep features (2048 dimensions) with a small CNN-trained feature set (128 dimensions) improved class separability. The rapid convergence of the CNN (loss decreasing from 0.0319 to 0.0207 across 3 epochs) illustrates its effectiveness as a feature enhancer.

- **Objective 3: Understanding class-specific performance.** Per-class evaluation showed consistently high F1-scores (0.94–0.99) for most categories. The highest performance was observed in *Forest*, *Residential*, and *SeaLake*, while comparatively lower recall was observed for *Highway* and *River*, likely due to visual ambiguity and structural similarity with other classes.

- **Unexpected Findings.** The Stacked Logistic Regression classifier did not significantly outperform XGBoost, indicating that deep features alone were highly discriminative, reducing the added benefit of meta-level stacking.

6.3 Contributions of the Research

6.3.1 Theoretical Contributions

This research contributes to the theoretical understanding of feature fusion in remote sensing classification. The study demonstrates that combining deep CNN features with lightweight CNN embeddings enhances class separation and improves classical model performance. It also reinforces findings from related literature that transfer learning from ImageNet-trained CNNs is highly effective for small- to medium-sized satellite datasets. [9]

6.3.2 Methodological Contributions

Methodologically, the study offers several contributions:

- A hybrid deep feature extraction pipeline integrating ResNet50 and a lightweight CNN.

- A fusion-based approach enabling classical classifiers such as XGBoost and Random Forest to achieve accuracy comparable to end-to-end deep learning models.

- A stacked meta-classifier framework that integrates probabilistic outputs from base models for improved prediction stability.

These techniques provide computationally efficient alternatives to fully fine-tuned deep networks, especially beneficial for systems with limited GPU resources.

6.3.3 Practical Implications

The findings have significant practical implications in domains such as:

- **Agriculture**: Monitoring crop types and vegetation levels.
- **Urban Planning**: Identifying residential, industrial, and transportation infrastructure.
- **Environmental Monitoring**: Segmenting water bodies, forests, and natural landscapes.

The hybrid framework is lightweight, adaptable, and suitable for deployment in real-time or large-scale geospatial systems.

6.4 Limitations of the Study

While the research produced promising results, certain limitations were identified:

- **GPU Memory Constraints**: The NVIDIA GTX 1650 (4 GB VRAM) limited batch sizes and model selection, restricting the use of larger architectures such as EfficientNet-B7 or ConvNeXt-Large.
- **Dataset Constraints**: The EuroSAT dataset contains 64x64 pixel images, which may limit spatial contextual learning. Larger-resolution datasets could yield more robust results.
- **Generalizability**: The model was evaluated only on EuroSAT RGB images. Its performance may differ on multispectral or hyperspectral datasets.
- **Class Confusion**: Similar land-cover classes (e.g., *Highway*, *Industrial*) introduced ambiguity, reducing accuracy for certain categories.

These limitations help contextualize the results and highlight areas for improvement.

6.5 Future Directions

6.5.1 Recommendations for Practice

Based on the findings, the following practical recommendations are proposed:

- Deploy the hybrid feature extraction model in operational geospatial systems where computational resources are limited, as it offers high accuracy with manageable processing requirements.

- Integrate additional geospatial features such as spectral indices (NDVI, SAVI) or temporal data to improve model discrimination for visually similar classes.
- Combine the model with GIS platforms for automated land management and environmental monitoring tasks.

6.5.2 Recommendations for Future Research

Future work may explore the following research directions:

- **End-to-End Fine-Tuning:** Implement transformers (e.g., ViT, Swin) or advanced CNNs (EfficientNetV2) to compare with the hybrid feature approach.
- **Multispectral and Hyperspectral Data:** Extend experiments to Sentinel-2 multispectral bands rather than RGB-only datasets.
- **Ensemble Learning:** Investigate more complex stacking mechanisms, including meta-learners such as Gradient Boosting or Neural Networks.
- **Spatial Context Modeling:** Introduce patch-based or attention-based architectures to incorporate surrounding spatial information.
- **Real-Time Deployment:** Evaluate performance on edge devices or cloud-based geospatial platforms.

These directions can significantly enhance the robustness and applicability of land-cover classification models.

6.6 Final Remarks

In conclusion, this research successfully demonstrated that hybrid deep feature extraction combined with classical machine learning classifiers can achieve high performance for land-cover classification using EuroSAT imagery. The findings contribute to both theoretical advancements and practical applications in remote sensing. It is hoped that this work will serve as a foundation for future innovations in geospatial intelligence, environmental monitoring, and satellite-based land analysis.

Bibliography

[1] Patrick Helber, Benjamin Bischke, Andreas Dengel, and Damian Borth. Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(7):2217–2226, 2019.

[2] Patrick Helber, Benjamin Bischke, Andreas Dengel, and Damian Borth. Introducing eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. In *IGARSS 2018-2018 IEEE international geoscience and remote sensing symposium*, pages 204–207. IEEE, 2018.

[3] Mohammad Alikhani, Behnam Ebadati, and Reza Attarzadeh. Lightweight and efficient deep learning models for remote sensing-based land use and land cover classification: A case study on eurosat dataset. *Earth Observation and Geomatics Engineering*, 8(2), 2024.

[4] Abhishek Bhatt and Vandana Thakur Bhatt. Dcrff-lhrf: an improvised methodology for efficient land-cover classification on eurosat dataset. *Multimedia Tools and Applications*, 83(18):54001–54025, 2024.

[5] Vladyslav Yaloveha, Andrii Podorozhniak, Heorhii Kuchuk, and Nataliia Garashchuk. Performance comparison of cnns on high-resolution multispectral dataset applied to land cover classification problem. *Radioelectronic and computer systems*, (2):107–118, 2023.

[6] Buse Saricayir and Caner Ozcan. Efficientnet deep learning model for satellite image classification using the eurosat dataset. 2025.

[7] Suman Kunwar and Jannatul Ferdush. Mapping of land use and land cover (lulc) using eurosat and transfer learning. *arXiv preprint arXiv:2401.02424*, 2023.

[8] Antonio Rangel, Juan Terven, Diana M Cordova-Esparza, and Edgar A Chavez-Urbiola. Land cover image classification. *arXiv preprint arXiv:2401.09607*, 2024.

[9] H Yassine, K Tout, and M Jaber. Improving lulc classification from satellite imagery using deep learning–eurosat dataset. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43:369–376, 2021.

[10] Agilandeeswari Loganathan, Suri Koushmitha, and Yerru Nanda Krishna Arun. Land use/land cover classification using machine learning and deep learning algorithms for eurosat dataset–a review. In *International Conference on Intelligent Systems Design and Applications*, pages 1363–1374. Springer, 2021.