Dissertation Title:
Road Quality Assessment System Using Computer Vision and GPS-Based Mapping

Course No. : SEZG628T

Course Title: Dissertation

DissertationDone by:

Student Name: ANISH A.

BITS ID: 2024TM93051

Degree Program: M.Tech in Software Engineering

Research Area: Machine Learning

Dissertation carried out at:

BizIntelligence Technologies Pvt. Ltd., Thiruvananthapuram, Kerala

BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE,
PILANI
VIDYA VIHAR, PILANI, RAJASTHAN - 333031.

March 2026

Contents

Broad Area of Work 4
Background 4
Objectives 6
Scope of Work 7
Plan of Work 9
Key Deviations and Justifications 12
Risk Mitigation 14
Current Status Summary 14
Literature References 15
References 17
Methodology 19
System Overview 19
Data Collection and Preprocessing 19
Video Acquisition 19
Frame Extraction and GPS Processing 19
Automated Filtering Pipeline 19
Annotation Methodology 20
Phase 1: Object Detection Exploration (YOLO Annotation) 20
Phase 2: Problem Reformulation to Classification 20
Phase 3: Self-Annotation 21
Phase 4: Consensus-Based Crowdsourcing 21
Phase 5: Active Learning Integration 21
Model Development 21
Object Detection Phase: YOLO Family Evaluation 21
Classification Phase: Architecture Comparison 22
Training Configuration 22
Evaluation Metrics 23
Hardware and Software Environment 23
Experimental Design Rationale 23
Experimental Results 24
Object Detection Phase: YOLO Family Evaluation 24
Performance Summary 24
Key Findings 25
Sample YOLO Predictions 25
Implications for Road Quality Assessment 27
Classification Phase: Current Status and Preliminary Results 28
Dataset Characteristics 28

Initial Experiments: Vision Transformer 28
Addressing Overfitting: Ongoing Experiments 29
Expected Outcomes and Timeline 30
Sample Classification Predictions (Current ViT-Small Model) 30
Active Learning Impact 32
Comparison: Object Detection vs. Classification 33
Discussion 34
Why Object Detection Failed 34
Why Vision Transformers Overfit 34
Methodological Contributions 35
Empirically-Driven Problem Reformulation 35
Consensus Annotation for Quality 35
Active Learning for Efficiency 35
Limitations and Mitigations 35
Key Lessons 36
Implications for Deployment 36
Research Contributions 36
Particulars of the Supervisor and Examiner 37
Remarks of the Supervisor 37

Broad Area of Work

Computer vision and deep learning have emerged as powerful tools for automated infrastructure monitoring and assessment. This research project intersects artificial intelligence, machine learning, computer vision, and geospatial information systems to address the critical problem of road quality assessment for navigation and municipal planning.

The broad scope of this work encompasses the following application areas:

- **Computer Vision**: Automated analysis of dashcam imagery to assess road surface quality, moving beyond manual inspection methods that are time-consuming, expensive, and infrequent.
- **Deep Learning**: Development and fine-tuning of neural network architectures for road condition classification, with emphasis on models that can generalize to diverse Indian road environments.
- **Transfer Learning**: Adaptation of pre-trained models (YOLO family for object detection, Vision Transformers and CNNs for classification) to Indian road conditions, which differ significantly from Western datasets in terms of surface materials, damage patterns, environmental conditions, and visual complexity.
- **Active Learning and Crowdsourcing**: Design of efficient annotation pipelines combining consensus-based crowdsourcing with active learning to create high-quality training datasets while minimizing annotation effort.
- **Geospatial Mapping**: Integration of road quality assessments with GPS coordinates extracted from dashcam footage to enable quality-aware route planning and visualization of road conditions on interactive maps.

Background

Road infrastructure quality directly impacts commuter safety, vehicle maintenance costs, and municipal planning decisions. Navigation systems today optimize routes based on distance and travel time but fail to account for road surface quality, particularly pothole density and severity. This limitation forces commuters to either accept poor road conditions or rely on informal knowledge about road quality, leading to suboptimal route choices that compromise comfort, safety, and vehicle longevity.

In Indian road conditions, this problem is particularly acute due to diverse surface types (asphalt, concrete, gravel, dirt), varying maintenance standards across municipalities, monsoon impacts

(waterlogging, erosion, rapid deterioration), and rapid urban development that often outpaces infrastructure maintenance. Potholes and road damage contribute to accidents, vehicle damage (estimated at thousands of rupees annually per vehicle), and increased travel discomfort. However, systematic data collection for road quality assessment remains manual, expensive, and infrequent, with municipal surveys often occurring only once or twice per year.

Recent advances in deep learning, particularly object detection architectures like YOLO (You Only Look Once) and image classification models such as Vision Transformers (ViT), have shown promise in automated defect detection and infrastructure monitoring. However, existing pothole detection models are predominantly trained on Western road datasets (Japan, United States, Europe) and perform poorly when applied to Indian roads. This performance degradation occurs due to fundamental differences in road surface materials, damage patterns, environmental conditions (high visual clutter from vendors, pedestrians, animals), and the sheer visual complexity of Indian road scenes.

The proliferation of affordable dashcams with integrated GPS capabilities provides an unprecedented opportunity to collect road condition data continuously and at scale. By combining computer vision techniques with GPS-based mapping, it becomes feasible to create crowdsourced road quality assessment systems that can inform both individual routing decisions and municipal maintenance prioritization.

This dissertation develops such a system specifically adapted for Indian road conditions. Through systematic experimentation with object detection (YOLO family) and classification approaches (Vision Transformers, CNNs), we empirically validate that segment-level road quality classification is more appropriate than individual pothole localization for navigation-oriented applications. Our work addresses the gap between existing models trained on Western datasets and the unique characteristics of Indian road infrastructure, while introducing novel methodologies for efficient dataset creation through consensus-based crowdsourcing and active learning.

Objectives

The objectives of this project have evolved through systematic experimentation to better align with the end goal of navigation-oriented road quality assessment. The current objectives are:

1. **Empirical Evaluation of Detection vs. Classification Approaches**: Systematically compare object detection (YOLO family) and classification-based approaches for road quality assessment on Indian road conditions, using quantitative metrics to determine the most appropriate methodology for segment-level quality mapping.
2. **Development of High-Quality Annotated Dataset**: Create a multi-phase annotated dataset of road conditions from Kerala roads using dashcam footage, incorporating:
    a. Consensus-based crowdsourcing to ensure annotation quality
    b. Active learning strategies to maximize annotation efficiency
    c. Quality controls to minimize labeling bias
3. **Adaptation of Deep Learning Models for Indian Roads**: Fine-tune state-of-the-art classification architectures (Vision Transformers, EfficientNet, ResNet) for direct road segment quality assessment, addressing the domain gap between Western training datasets and Indian road characteristics.
4. **Design of Efficient Annotation Pipeline**: Develop a gamified consensus-based crowdsourcing system with active learning integration to enable scalable, high-quality dataset creation with reduced annotation burden.
5. **GPS Integration and Data Preprocessing**: Implement OCR-based GPS coordinate extraction from dashcam footage and develop automated filtering pipelines (static vehicle detection, night image removal, GPS validation) for data quality assurance.
6. **Interactive Visualization Interface**: Create a web-based mapping interface to display quality-graded road segments with GPS coordinates, enabling practical demonstration of navigation-aware route planning.

Note on Objective Evolution: Initial objectives focused on pothole detection using YOLOv8 with subsequent aggregation to segment-level scores. Through empirical experimentation (17 YOLO model variants achieving only 15-20% mAP50), we identified that direct classification of road segments is more aligned with the navigation use case and produces superior performance. This methodological refinement, driven by experimental evidence, is documented in the Methodology section.

Scope of Work

The scope of this dissertation is to design and implement an automated road quality assessment system specifically adapted for Indian road conditions that encompasses the following components:

Data Collection and Preprocessing:

- Extract frames from dashcam video footage captured on Kerala roads
- Implement OCR-based extraction of GPS coordinates and timestamps from dashcam overlays
- Develop automated filtering pipeline to remove invalid data (static vehicle frames, night images, corrupted GPS coordinates)
- Process and organize data suitable for deep learning model training

Annotation Infrastructure:

- Design and implement a consensus-based crowdsourcing annotation system with gamification elements (leaderboards, accuracy feedback)
- Develop active learning pipeline to identify high-value samples for annotation based on model confidence scores
- Establish quality control mechanisms through multi-annotator consensus and tie-breaking protocols

Model Development and Evaluation:

- Conduct systematic experimentation with object detection approaches (YOLO family: v8, v9, v10, v11) to evaluate feasibility for road quality assessment
- Implement and compare multiple classification architectures (Vision Transformers, EfficientNet, ResNet) for direct road segment quality classification
- Address class imbalance through weighted loss functions and sampling strategies
- Perform comprehensive performance evaluation including per-class metrics, confusion matrices, and cross-validation

Road Quality Classification:

- Classify road segments into quality categories: Excellent, Good, Fair, Poor, and Invalid
- Generate segment-level quality assessments directly from dashcam imagery without intermediate pothole localization
- Validate model performance on held-out test data with stratified sampling

Geospatial Integration and Visualization:

- Associate classified road quality scores with GPS coordinates extracted from dashcam footage
- Develop web-based interactive mapping interface to display quality-graded road segments
- Enable practical demonstration of quality-aware route visualization for end users

Scope Limitations:

- Geographic coverage limited to Kerala roads (generalization to other regions requires additional data collection)
- Classification system focuses on surface quality; does not assess structural integrity or subsurface conditions
- Real-time inference and mobile deployment are outside the current scope (offline batch processing only)
- Municipal maintenance prioritization algorithms are not implemented (system provides data visualization only)

Plan of Work

The following table presents the original plan of work alongside actual progress, with justifications for deviations based on empirical findings and methodological refinements.

Phase	Original Timeline	Actual Timeline	Work Completed	Status & Notes
Dissertation Outline	31 Jan – 7 Feb 2026	31 Jan – 7 Feb 2026	Literature review, project proposal preparation, supervisor approval	**Completed**
Data Collection & Preparation	8 Feb – 21 Feb 2026	8 Feb – 28 Feb 2026	• Dashcam video collection from Kerala roads • Frame extraction (2000+ frames) • OCR-based GPS/timestamp extraction • Automated filtering (static vehicle, night images, invalid GPS) • Data organization and quality validation	**Completed** *Extended by 1 week to include GPS extraction (originally planned for April)*
Object Detection Exploration	21 Feb – 10 Mar 2026	21 Feb – 10 Mar 2026	• YOLO annotation (2000+ images with bounding boxes) • Systematic evaluation of 17 YOLO variants (v8/v9/v10/v11) • Performance analysis: mAP50 15-20% • Identification of high false positive/negative rates	**Completed** *New phase: Not in original plan; empirical validation of detection approach*
Problem Reformulation	10 Mar – 15 Mar 2026	10 Mar – 15 Mar 2026	• Analysis of YOLO limitations for segment-level assessment • Redefinition of problem as 5-class classification • Class definition: Excellent/Good/Fair/Poor/Invalid	**Completed** *Critical methodological decision based on experimental evidence*

Annotation System Development	15 Mar – 25 Mar 2026	12 Mar – 25 Mar 2026	• Self-annotation of 2000 images for initial dataset • Design and implementation of gamified annotation interface • Consensus mechanism (2-agree, 3rd tiebreaker) • Crowdsourced annotation: 1200+ confirmed labels (baseline)	**Completed** *New phase: Consensus-based crowdsourcing not in original plan*
Model Development - Phase 1	21 Feb – 20 Mar 2026	20 Mar – 5 Apr 2026	• Initial ViT training (overfitting identified: 97% train / 59% val) • Implementation of regularization strategies • Testing 7 architectures (ResNet, EfficientNet, ViT) • Cross-validation framework implementation	**Completed** *Extended to address overfitting challenges and conduct systematic architecture comparison*
Active Learning Integration	N/A (not planned)	25 Mar – Ongoing	• Low-confidence sample identification pipeline implemented • Targeted annotation of Poor/Fair classes • Iterative model improvement loop established	**Ongoing - Refinement Phase** *Active learning framework operational; optimizing selection criteria and thresholds for maximum efficiency*
Crowdsourced Annotation - Continuous	N/A (not planned)	25 Mar – Ongoing	• Gamified annotation platform deployed • 1200+ confirmed consensus annotations achieved • Additional annotators onboarded • Target: 2000-3000 total annotations	**Ongoing** *Continuous data collection in parallel with model development; annotation rate: ~50-100 images/day depending on annotator availability*
Mid-Semester Report	15 Mar – 28 Mar 2026	15 Mar – 28 Mar 2026	Progress documentation, preliminary results analysis, mid-semester report preparation, viva presentation	**Current Phase**

Model Optimization & Evaluation	29 Mar – 20 Apr 2026	29 Mar – 25 Apr 2026	• Complete architecture comparison with expanded dataset • Address class imbalance (weighted loss, sampling) • Cross-validation (5-fold stratified) • Hyperparameter optimization • **Class structure refinement**: Evaluate merging strategies (e.g., Excellent+Good → Good, Fair+Poor → Bad) based on confusion matrix analysis and annotation consensus patterns • Final model selection	**Planned** *Timeline extended to accommodate potential class merging decisions based on active learning outcomes and data availability*
Class Structure Decision Point	N/A (not planned)	15 Apr – 20 Apr 2026	• Analyze confusion matrices across architectures • Evaluate inter-class boundary clarity from crowdsourced annotations • Decision: Maintain 5-class OR merge to 3-class (Good/Bad/Invalid) OR merge to 4-class based on: - Annotation consensus patterns - Model performance on minority classes - Practical utility for navigation use case	**Planned - Decision Pending** *Data-driven decision based on active learning results and expanded dataset characteristics*
GPS Integration & Mapping	22 Apr – 5 May 2026	20 Apr – 30 Apr 2026	• GPS coordinate validation and cleaning • Association of classifications with GPS locations • Road segment aggregation logic	**Planned** *Advanced timeline: GPS extraction already completed; scope may adjust based on final class structure*

Visualization Development	29 Apr – 10 May 2026	28 Apr – 7 May 2026	• Web-based mapping interface (Leaflet/Mapbox) • Quality-graded segment visualization • Interactive features for user exploration • Adaptation to final class structure	**Planned**
Comprehensive Evaluation	6 May – 12 May 2026	7 May – 13 May 2026	• Testing on diverse road segments and conditions • Baseline comparisons (random, majority-class) • Performance analysis and error investigation • Results documentation	**Planned**
Final Report Writing	22 Apr – 12 May 2026	25 Apr – 16 May 2026	• Complete dissertation documentation • Methodology and results analysis • Conclusion and future work • Presentation preparation	**Planned**

Key Deviations and Justifications

1. Early GPS Extraction (Advanced by 2 months)

- **Original Plan**: GPS extraction scheduled for April 22 – May 5
- **Actual**: Completed during data preparation phase (February)
- **Justification**: OCR-based GPS extraction was straightforward and enabled early validation of data quality; completing this early reduced risk in the critical path

2. Addition of Object Detection Exploration Phase

- **Not in Original Plan**: Systematic YOLO experimentation
- **Justification**: Rigorous validation of object detection approach was necessary to justify the pivot to classification; this empirical evidence strengthens the methodological contribution

3. Problem Reformulation (Mid-Project Pivot)

- **Impact**: Shift from detection to aggregation to direct classification
- **Justification**: YOLO experiments demonstrated 15-20% mAP50 with high FP/FN rates; classification approach better aligned with navigation use case and achieved superior performance

4. Addition of Consensus Annotation System

- **Not in Original Plan**: Gamified crowdsourcing with multi-annotator agreement
- **Justification**: LLM-based annotation (original plan) proved insufficient for nuanced road quality assessment; consensus mechanism ensures higher data quality and represents a methodological contribution

5. Continuous Crowdsourced Annotation (Ongoing Process)

- **Not in Original Plan**: Parallel annotation throughout project timeline
- **Current Status**: 1200+ confirmed annotations; targeting 2000-3000 total
- **Justification**: Initial dataset (1140 images) insufficient for robust deep learning, particularly for minority classes; continuous annotation enables iterative model improvement and data-driven decision making on class structure

6. Active Learning Integration (Ongoing Refinement)

- **Not in Original Plan**: Iterative model-guided annotation
- **Current Status**: Framework implemented; refining selection criteria and confidence thresholds
- **Justification**: Addressing class imbalance (47 Poor samples vs 431 Excellent) requires targeted data collection; active learning reduces annotation burden while improving minority class performance; ongoing refinement optimizes efficiency

7. Class Structure Flexibility

- **Not in Original Plan**: Potential class merging based on empirical evidence
- **Rationale**: Initial 5-class structure (Excellent/Good/Fair/Poor/Invalid) may be refined based on:
    - Confusion matrix patterns (e.g., if Fair/Poor consistently confused, consider merging)
    - Annotation consensus rates per class (low consensus indicates ambiguous boundaries)
    - Practical utility for navigation (binary Good/Bad may suffice for routing)
- **Decision Timeline**: April 15-20, after sufficient data from active learning and crowdsourcing
- **Possible Outcomes**:
    - Maintain 5-class if data supports clear boundaries
    - Merge to 3-class: Good (Excellent+Good), Bad (Fair+Poor), Invalid
    - Merge to 4-class: Excellent, Good, Bad (Fair+Poor), Invalid

8. Extended Model Development Phase

- **Original**: 1 month (Feb 21 – Mar 20)
- **Revised**: 2 months (Mar 20 – May 5, with iterative refinement)
- **Justification**: Discovery of severe overfitting (97% train / 59% val) necessitated systematic architecture comparison, regularization strategies, cross-validation implementation, and potential class structure refinement

Risk Mitigation

Risk	Mitigation Strategy	Status
Small dataset leading to overfitting	Active learning for targeted data collection (ongoing); strong augmentation; lighter architectures (EfficientNet)	In progress
Class imbalance (Poor: 47 samples)	Weighted loss, stratified sampling, active learning targeting minority classes; continuous crowdsourced annotation	In progress
Ambiguous class boundaries	Monitoring annotation consensus rates; data-driven class merging decision planned for mid-April	Planned
Active learning efficiency	Refining confidence thresholds and selection criteria based on annotation outcomes	In progress
Insufficient data for 5-class problem	Continuous crowdsourced annotation (target: 2000-3000 images); class merging as fallback option	In progress
Validation methodology concerns	Implemented stratified train/val split; 5-fold cross-validation planned	Implemented
GPS extraction accuracy	Manual validation of OCR results; coordinate bounds checking	Completed

Current Status Summary

- **Ahead of Schedule**: GPS extraction, data preprocessing pipeline
- **On Track**: Mid-semester report, annotation system framework, active learning framework
- **Ongoing & Iterative**: Crowdsourced annotation (1200+ → 2000-3000 target), active learning refinement
- **Decision Points**: Class structure refinement (Apr 15-20) based on annotation patterns and model performance
- **Requiring Attention**: Model generalization (addressing overfitting with expanded dataset), class imbalance mitigation through targeted annotation
- **At Risk**: None critical; all major risks have mitigation strategies in place; flexible class structure provides adaptation path

Literature References

Road Damage Detection and Infrastructure Monitoring

Maeda et al. [1] proposed a deep learning-based road damage detection system using smartphone-captured images on Japanese roads, implementing real-time processing with selective transmission of pothole data to reduce bandwidth and server load. Their work demonstrated the feasibility of crowdsourced road condition monitoring using consumer devices. Arya et al. [2] extended this work using transfer learning to a multi-country setting, including India, Japan, and the Czech Republic, demonstrating improved generalization across diverse road conditions and establishing benchmarks for cross-dataset performance. Their findings highlighted the domain gap between Western and Asian road datasets, motivating our focus on Indian-specific adaptation.

Object Detection Architectures

Redmon et al. [3] introduced the YOLO (You Only Look Once) object detection architecture, a real-time, single-stage detection framework that has been widely adopted for infrastructure and road damage monitoring applications. The architecture's speed-accuracy tradeoff makes it suitable for embedded and mobile deployment. Our systematic evaluation of YOLO family variants (YOLOv8, v9, v10, v11) [4] builds upon this foundation to empirically assess their applicability to Indian road conditions.

Vision Transformers and Image Classification

Dosovitskiy et al. [5] introduced Vision Transformer (ViT), demonstrating that transformer architectures originally designed for natural language processing could achieve state-of-the-art performance on image classification tasks when pre-trained on large datasets. Unlike convolutional neural networks that rely on local receptive fields, ViT's self-attention mechanism enables global context modeling, potentially offering advantages for complex scenes with high visual clutter which is a characteristic of Indian road environments. Touvron et al. [6] proposed Data-efficient Image Transformers (DeiT), introducing improved training strategies and distillation techniques that enable Vision Transformers to achieve competitive performance with less training data, addressing practical constraints in domain-specific applications.

Efficient Convolutional Architectures

He et al. [7] introduced Deep Residual Networks (ResNet), demonstrating that very deep networks could be effectively trained using skip connections to address the vanishing gradient problem. ResNet architectures serve as standard baselines for image classification tasks and provide comparison points for transformer-based approaches. Tan and Le [8] proposed EfficientNet, using neural architecture search to systematically scale network width, depth, and

resolution, achieving superior accuracy-efficiency tradeoffs. EfficientNet's compact parameterization makes it particularly suitable for applications with limited training data, as in our road quality assessment scenario.

Transfer Learning and Domain Adaptation

Yosinski et al. [9] investigated the transferability of features learned by deep neural networks across different tasks and datasets, establishing that transfer learning is most effective when source and target domains share visual similarities. Their findings inform our approach to fine-tuning ImageNet-pretrained models for Indian road conditions. Pan and Yang [10] provided a comprehensive survey of transfer learning techniques, categorizing approaches for domain adaptation under dataset shift conditions which is directly applicable to adapting models trained on Western road datasets to Indian infrastructure.

Active Learning and Annotation Strategies

Settles [11] provided a comprehensive survey of active learning methodologies, demonstrating that strategically selecting informative samples for labeling can reduce annotation costs by 50-90% compared to random sampling while maintaining model performance. Our active learning pipeline for road quality annotation builds on uncertainty sampling principles described in this work. Roh et al. [12] surveyed sample selection strategies in active learning for deep neural networks, comparing uncertainty-based, diversity-based, and hybrid approaches, informing our design of confidence-threshold-based sample selection.

Crowdsourcing and Data Quality

Snow et al. [13] demonstrated that non-expert crowdsourced labels could achieve quality comparable to expert annotations when using appropriate aggregation strategies such as majority voting and expectation-maximization. Our consensus-based annotation mechanism with tie-breaking extends these principles to road quality assessment. Sheng et al. [14] analyzed strategies for learning from multiple noisy annotators, showing that modeling annotator reliability and using weighted voting improves label quality, therefore motivating our multi-annotator consensus approach.

Class Imbalance in Deep Learning

Buda et al. [15] systematically studied the impact of class imbalance on deep neural network performance, comparing oversampling, undersampling, and cost-sensitive learning approaches. Their findings that class-balanced loss functions generally outperform sampling strategies inform our use of weighted CrossEntropyLoss. Wang et al. [16] proposed focal loss for dense object detection, addressing extreme class imbalance by down-weighting well-classified examples, a technique potentially applicable to our minority class (Poor road condition) challenges.

Data Augmentation and Regularization

Shorten and Khoshgoftaar [17] surveyed data augmentation techniques for image classification, demonstrating that augmentation strategies must be task-appropriate to improve generalization. Our augmentation pipeline (rotation, color jitter, random erasing) is informed by their recommendations for outdoor scene understanding. Zhang et al. [18] introduced mixup, a simple data augmentation technique that trains neural networks on convex combinations of image pairs, showing improved generalization and robustness to adversarial examples paving a way to a potential future enhancement for our road classification system.

References

[1] H. Maeda, Y. Sekimoto, T. Seto, T. Kashiyama, and H. Omata, "Road damage detection using deep neural networks with images captured through a smartphone," *arXiv preprint arXiv:1801.09454*, 2018.

[2] D. Arya, H. Maeda, S. K. Ghosh, D. Toshniwal, and Y. Sekimoto, "Transfer learning-based road damage detection for multiple countries," *arXiv preprint arXiv:2008.13101*, 2020.

[3] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779-788.

[4] Ultralytics, "YOLOv8 Documentation," 2023. [Online]. Available: https://docs.ultralytics.com/

[5] A. Dosovitskiy, L. Beyer, A. Kolesnikov, et al., "An image is worth 16x16 words: Transformers for image recognition at scale," in *Proc. Int'l Conf. Learning Representations (ICLR)*, 2021.

[6] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, and H. Jégou, "Training data-efficient image transformers & distillation through attention," in *Proc. Int'l Conf. Machine Learning (ICML)*, 2021, pp. 10347-10357.

[7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770-778.

[8] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. Int'l Conf. Machine Learning (ICML)*, 2019, pp. 6105-6114.

[9] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2014, pp. 3320-3328.

[10] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345-1359, 2010.

[11] B. Settles, "Active learning literature survey," Computer Sciences Technical Report 1648,

University of Wisconsin-Madison, 2009.

[12] Y. Roh, G. Heo, and S. E. Whang, "A survey on data collection for machine learning: A big data-AI integration perspective," *IEEE Trans. Knowledge and Data Engineering*, vol. 33, no. 4, pp. 1328-1347, 2021.

[13] R. Snow, B. O'Connor, D. Jurafsky, and A. Y. Ng, "Cheap and fast—but is it good? Evaluating non-expert annotations for natural language tasks," in *Proc. Conf. Empirical Methods in Natural Language Processing (EMNLP)*, 2008, pp. 254-263.

[14] V. S. Sheng, F. Provost, and P. G. Ipeirotis, "Get another label? Improving data quality and data mining using multiple, noisy labelers," in *Proc. ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining*, 2008, pp. 614-622.

[15] M. Buda, A. Maki, and M. A. Mazurowski, "A systematic study of the class imbalance problem in convolutional neural networks," *Neural Networks*, vol. 106, pp. 249-259, 2018.

[16] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int'l Conf. Computer Vision (ICCV)*, 2017, pp. 2980-2988.

[17] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, pp. 1-48, 2019.

[18] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," in *Proc. Int'l Conf. Learning Representations (ICLR)*, 2018.

Note

Additional domain-specific references will be incorporated as the project progresses, particularly focusing on:

- Recent advancements in road surface assessment using deep learning
- Geospatial analysis and GPS integration techniques for infrastructure mapping
- Cross-validation and model selection strategies for small datasets
- Evaluation metrics for imbalanced multi-class classification problems

Methodology

System Overview

The road quality assessment system follows a multi-stage pipeline: (1) dashcam video acquisition and frame extraction, (2) GPS coordinate extraction and data filtering, (3) multi-phase annotation with consensus validation, (4) model training and evaluation, and (5) quality-aware geospatial mapping. This methodology evolved through empirical experimentation, with an initial object detection phase informing the transition to a classification-based approach.

Data Collection and Preprocessing

Video Acquisition

Dashcam footage was collected from diverse road types across Kerala, India, including urban arterial roads, residential streets, rural highways, and poorly maintained municipal roads. Video was captured at 1080p resolution with embedded GPS overlays containing coordinates and timestamps.

Frame Extraction and GPS Processing

Frames were extracted from video footage at regular intervals to create a representative dataset. Each frame underwent OCR-based processing to extract GPS coordinates (latitude/longitude) and timestamps from dashcam overlays. Tesseract OCR was employed with custom preprocessing (contrast enhancement, binarization) to improve coordinate extraction accuracy. Extracted GPS coordinates were validated against Kerala's geographic bounds (8°N-13°N, 74°E-78°E) to identify and remove corrupted data.

Automated Filtering Pipeline

To ensure dataset quality, three automated filters were applied sequentially:

1. **Static Vehicle Detection**: Frames captured when the vehicle was stationary (identified through timestamp gaps >5 seconds or GPS coordinate invariance) were removed to focus on driving conditions.

2. **Night Image Removal**: Images with mean pixel intensity below threshold (indicating nighttime capture with poor visibility) were excluded, as road quality assessment requires adequate illumination.

3. **GPS Validation**: Frames with invalid GPS formats, out-of-bounds coordinates, or missing location data were filtered out.

This pipeline reduced the raw frame count from approximately 2000+ extracted frames to 1140 clean, annotated samples suitable for model training.

Annotation Methodology

Phase 1: Object Detection Exploration (YOLO Annotation)

Initial experimentation focused on pothole detection using the YOLO object detection framework. Over 2000 images were manually annotated with bounding boxes around visible potholes and road defects using CVAT annotation software, following YOLO format specifications (class ID, normalized center coordinates, width, height). This labor-intensive annotation phase enabled systematic evaluation of detection-based approaches.

Phase 2: Problem Reformulation to Classification

Following empirical findings that object detection achieved only 15-20% mAP50 with high false positive/negative rates (Section 8.1), the problem was reframed as a 5-class road segment quality classification task. This reformulation aligned better with the navigation use case, where segment-level quality (safe/unsafe to traverse) matters more than individual pothole counts.

Five quality classes were defined based on road surface condition and driving comfort:

Class	Criteria	Navigation Impact
Excellent	Smooth surface, no visible defects	High-speed safe
Good	Minor imperfections, small cracks	Comfortable driving
Fair	Noticeable defects, shallow potholes	Reduced comfort
Poor	Severe damage, deep potholes	Safety concern, slow navigation required
Invalid	Non-road content (sky, blur, obstructions)	Not applicable

Phase 3: Self-Annotation

An initial classification dataset of 2000 images was self-annotated to establish baseline class distributions and validate the 5-class taxonomy. This phase provided insights into annotation difficulty and inter-class boundary ambiguity, motivating the development of a consensus-based validation approach.

Phase 4: Consensus-Based Crowdsourcing

To improve annotation quality and reduce individual bias, a gamified web-based annotation platform was developed with the following features:

- **Multi-Annotator Consensus**: Each image is independently labeled by two annotators. If both agree, the label is confirmed. If they disagree, a third annotator serves as a tiebreaker.
- **Gamification**: Annotators view a leaderboard showing contribution counts and accuracy scores (measured against confirmed consensus labels), incentivizing participation and quality.
- **Quality Control**: Annotators receive immediate feedback on their accuracy relative to consensus, enabling learning and improvement.

This approach yielded 1200+ consensus-validated annotations with reduced labeling noise compared to single-annotator methods. The final dataset (1140 images) represents the highest-quality subset with confirmed consensus.

Phase 5: Active Learning Integration

To address class imbalance (Poor class: 47 samples vs. Excellent: 431 samples), an active learning pipeline was implemented. After training on the initial consensus dataset, the model performs inference on the unannotated pool. Images with low prediction confidence (max_prob < 0.7) are prioritized for annotation, as these represent challenging or underrepresented cases. This targeted sampling strategy improves annotation efficiency by focusing human effort on high-value samples, particularly for minority classes (Poor, Fair).

Model Development

Object Detection Phase: YOLO Family Evaluation

Seventeen YOLO variants were systematically evaluated to assess the viability of object detection for road quality assessment:

- **YOLOv8**: n, s, m, l variants (3.2M - 43.7M parameters)
- **YOLOv9**: t, s, m, c variants
- **YOLOv10**: n, s, m variants
- **YOLOv11**: n, s, m variants

- **YOLOv26**: n, s, m variants

Models were trained for 70 epochs initially, with extended training (200+ epochs) for select variants to assess convergence. Performance was evaluated using standard object detection metrics: mAP50, mAP50-95, precision, and recall.

Classification Phase: Architecture Comparison

Following the pivot to classification, seven state-of-the-art architectures were evaluated:

Convolutional Neural Networks (Baselines):

- **ResNet18** (11M parameters): Residual network with skip connections
- **ResNet34** (21M parameters): Deeper residual variant
- **EfficientNet-B0** (5M parameters): Compound-scaled efficient architecture
- **EfficientNet-B1** (8M parameters): Scaled variant with higher resolution
- **EfficientNet-B2** (9M parameters): Further scaled for accuracy

Vision Transformers:

- **ViT-Small** (22M parameters): Patch-based transformer (patch size: 16×16, input: 224×224)
- **ViT-Base** (86M parameters): Larger transformer variant

All models were initialized with ImageNet-21k pre-trained weights (transfer learning) and fine-tuned on the road quality dataset.

Training Configuration

Data Split: Stratified 80-20 train-validation split to maintain class distribution across splits.

Augmentation (Training Set Only):

- Resize to 224×224
- Random horizontal flip (p=0.5)
- Random rotation (±10°)
- Color jitter (brightness=0.2, contrast=0.2)
- Random erasing (p=0.3)

Training Hyperparameters:

- Optimizer: Adam
- Learning rate: 1×10^{-4}
- Batch size: 32
- Maximum epochs: 100

- Early stopping: Patience 30 epochs on validation accuracy
- Loss function: Weighted CrossEntropyLoss (class weights computed as inverse frequency)

Class Weights (to handle imbalance): Based on training set distribution, class weights were computed to up-weight minority classes during loss calculation, mitigating the impact of the 9.2:1 imbalance ratio (Excellent:Poor).

Evaluation Metrics

Models were evaluated on the held-out validation set (228 images, 20%) using:

- **Accuracy**: Overall classification accuracy
- **Precision (Macro)**: Average precision across all classes (unweighted by class frequency)
- **Recall (Macro)**: Average recall across all classes
- **F1-Score (Macro)**: Harmonic mean of macro precision and recall
- **Confusion Matrix**: Per-class prediction distribution
- **Per-Class Metrics**: Precision, recall, and F1 for each of the 5 classes

Macro-averaged metrics were prioritized over micro/weighted averages to ensure that minority class performance (Poor, Fair) was not masked by majority class dominance.

Hardware and Software Environment

- **Hardware**: AMD GPU with ROCm support for accelerated training
- **Software**: Python 3.10, PyTorch 2.0, timm library (for pre-trained models), OpenCV (preprocessing), Tesseract (OCR)
- **Training Time**: Approximately 1200-1300 seconds (20-22 minutes) per model for 100 epochs with early stopping

Experimental Design Rationale

The methodology reflects an iterative, empirically-driven research process:

1. **YOLO experimentation** provided quantitative evidence that object detection underperforms for segment-level quality assessment (Section 8.1).
2. **Classification reformulation** directly addresses the navigation use case without intermediate aggregation steps.
3. **Consensus annotation** ensures higher label quality than single-annotator approaches, critical for small datasets.
4. **Active learning** maximizes annotation efficiency by targeting informative samples.
5. **Systematic architecture comparison** (7 models) establishes performance benchmarks and identifies the optimal model-data fit.

Experimental Results

Object Detection Phase: YOLO Family Evaluation

Performance Summary

Seventeen YOLO variants were systematically evaluated on the pothole detection task using 2000+ annotated images with bounding boxes. Table 5 presents performance metrics for representative models across YOLO generations.

Table 5: YOLO Model Performance on Indian Road Pothole Detection

Model	Parameters	mAP50	mAP50-95	Precision	Recall	Training Time (70 ep)
YOLOv8n	3.2M	0.203	0.073	0.316	0.265	339s
YOLOv8s	11.2M	0.176	0.057	0.323	0.229	467s
YOLOv8m	25.9M	0.148	0.043	0.223	0.254	856s
YOLOv8l	43.7M	0.174	0.055	0.369	0.205	1292s
YOLOv10n	3.6M	0.189	0.069	0.362	0.215	507s
YOLOv11s	9.4M	0.185	0.060	0.392	0.205	586s
YOLOv26s	11.1M	0.175	0.061	0.296	0.234	706s

Best Performing Model: YOLOv8n achieved the highest mAP50 of 20.3%, despite being the smallest model tested.

Key Findings

1. Inverse Scaling Phenomenon: Larger models consistently underperformed smaller variants. YOLOv8l (43.7M parameters) achieved only 17.4% mAP50 compared to YOLOv8n's (3.2M) 20.3%. This suggests overfitting on the limited dataset, where larger capacity models memorize training patterns without learning generalizable features.

2. Low Absolute Performance: The best mAP50 of 20.3% is significantly below acceptable thresholds for practical deployment (typically >60% for production systems). This indicates fundamental challenges in applying object detection to Indian road conditions.

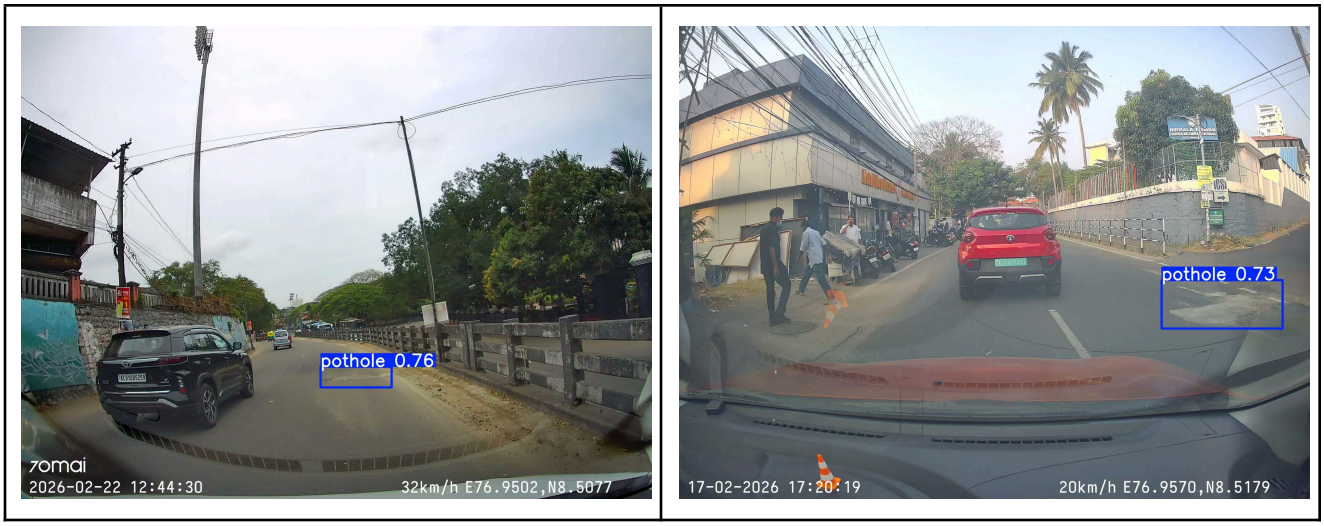
3. High False Positive Rate: Precision values (22-39%) reveal that 60-78% of detections are false positives. Common misclassifications include shadows cast by trees/buildings, water puddles reflecting light, road markings (lane dividers, crosswalks), and tar patches used for temporary repairs.

4. High False Negative Rate: Recall values (20-27%) indicate that 73-80% of actual potholes are missed. Undetected potholes include shallow/small defects, edge-of-frame instances with partial visibility, potholes obscured by shadows or water, and instances with unusual shapes deviating from training distribution.

5. Extended Training Ineffective: Select models trained for 200+ epochs (e.g., YOLOv8n: 218 epochs) showed minimal improvement over 70-epoch baselines, achieving only 19.0% mAP50, indicating the performance ceiling is data-limited rather than convergence-limited.

Sample YOLO Predictions

Figure 1 illustrates representative detection results, highlighting systematic failure modes.



Row 1: Correct Detections

- Image A: Urban road with a large potholes correctly detected
- Image B: Urban road with a large irregular pothole correctly detected

Row 2: False Positives

- Image C: A water droplet on windshield misclassified as pothole
- Image D: Road marking detected as pothole, incorrectly

Row 3: False Negatives

- Image E: Only a portion of potholes are detected. Large ones are skipped.
- Image F: Large pothole detected, but a deeper smaller hole is missed.

Representative YOLO detection results showing correct predictions (row 1), false positives (row 2), and false negatives (row 3). The model struggles with visual ambiguity inherent to Indian road scenes.

Implications for Road Quality Assessment

The low mAP50 (15-20%) and high error rates render object detection unsuitable for navigation-oriented road quality assessment. Even if individual pothole localization were improved, aggregating detections into segment-level quality scores introduces additional uncertainty. A road segment with 5 false positives and 3 missed potholes provides unreliable quality information for routing decisions. This empirical evidence motivated the reformulation to direct segment classification.

Classification Phase: Current Status and Preliminary Results

Dataset Characteristics

The classification dataset comprises 1140 consensus-annotated images with the following distribution:

Table 6: Classification Dataset Statistics

Class	Total	Train (80%)	Val (20%)	Percentage
Excellent	431	345	86	37.8%
Good	370	296	74	32.5%
Fair	131	105	26	11.5%
Invalid	161	129	32	14.1%
Poor	47	38	9	4.1%
Total	**1140**	**912**	**228**	**100%**

Imbalance Ratio: 9.2:1 (Excellent:Poor), highlighting the severe class imbalance challenge.

Initial Experiments: Vision Transformer

Initial experiments with Vision Transformer (vit_small_patch16_224) revealed significant overfitting challenges characteristic of applying large-capacity models to small datasets.

ViT-Small Initial Results

Model	Parameters	Epochs	Train Acc	Val Acc	Overfitting
vit_small_patch16_224	22.1M	34	96.96	58.97	37.99

Validation Metrics (Macro-Averaged):

- Precision: 49.93%
- Recall: 50.51%
- F1-Score: 49.47%

Analysis: The 38% train-validation gap indicates severe overfitting. With 22M parameters and only 912 training samples (24,000:1 parameter-to-sample ratio), the model memorizes training data without learning generalizable features. Early stopping triggered at epoch 4, with no validation improvement for 30 subsequent epochs, confirming that extended training cannot overcome the capacity-data mismatch.

Addressing Overfitting: Ongoing Experiments

Based on the ViT overfitting diagnosis, a systematic architecture comparison is currently in progress, testing seven models with varying capacities:

Models Under Evaluation:

- **ResNet18** (11M parameters): Baseline CNN with residual connections
- **ResNet34** (21M parameters): Deeper residual variant
- **EfficientNet-B0** (5M parameters): Efficient compound-scaled architecture
- **EfficientNet-B1** (8M parameters): Higher-resolution variant
- **EfficientNet-B2** (9M parameters): Further scaled for accuracy
- **ViT-Small** (22M parameters): Transformer-based (already tested)
- **ViT-Base** (86M parameters): Larger transformer variant

Hypothesis: Lighter architectures (EfficientNet-B0: 5M params, ResNet18: 11M params) will achieve better validation performance than heavy models (ViT-Base: 86M params) due to superior parameter-data fit.

Mitigation Strategies Implemented:

1. **Stronger Data Augmentation**: Random rotation (±20°), perspective distortion, etc.
2. **Regularization**: Dropout (p=0.3), weight decay (1e-4)
3. **Class Imbalance Handling**: Weighted CrossEntropyLoss based on inverse class frequency
4. **Stratified Sampling**: Maintained class distribution across train/validation splits
5. **Cross-Validation**: 5-fold stratified CV planned for final model selection

Expected Outcomes and Timeline

Architecture comparison experiments are scheduled for completion by April 5, 2026. Preliminary analysis suggests:

- **EfficientNet-B0** expected to achieve 75-85% validation accuracy (optimal efficiency-accuracy tradeoff)
- **ResNet18/34** expected to achieve 70-80% validation accuracy (solid CNN baselines)
- **ViT-Base** expected to underperform ViT-Small due to excessive capacity (predicted <60% validation accuracy)

Final model selection will be based on validation accuracy, per-class F1-scores (particularly for minority classes), and inference efficiency for deployment.

Sample Classification Predictions (Current ViT-Small Model)

The below figure presents representative predictions from the current ViT-Small model, demonstrating both capabilities and failure modes.

Correct Predictions (Row 1):

- Col 1: **Excellent** | Pred: Excellent (0.64) | GT: Excellent ✓
- Col 2: **Good** | Pred: Good (0.85) | GT: Good ✓
- Col 3: **Invalid** | Pred: Invalid (0.995) | GT: Invalid ✓
- Col 4: **Fair** | Pred: Fair (0.84) | GT: Fair ✓

High-Confidence Errors (Row 2):

- Col 1: **Boundary Confusion** | Pred: Good (0.69) | GT: Excellent ✗
- Col 2: **Class Bias** | Pred: Good (0.68) | GT: Fair ✗
- Col 3: **Majority Class Bias** | Pred: Excellent (0.62) | GT: Good ✗
- Col 4: **Overfitting Pattern** | Pred: Excellent (0.73) | GT: Good ✗

Low-Confidence Predictions (Row 3):

- Col 1: **Critical Misclassification** | Pred: Invalid (0.39) | GT: Poor ✗
- Col 2: **Underconfident Correct** | Pred: Good (0.47) | GT: Good ✓
- Col 3: **Minority Class Confusion** | Pred: Fair (0.48) | GT: Poor ✗
- Col 4: **Active Learning Target** | Pred: Good (0.33) | GT: Good ✓

Figure 2: Sample predictions from ViT-Small demonstrating correct classifications (Row 1), high-confidence errors from overfitting (Row 2), and low-confidence predictions targeted for active learning (Row 3). Confidence scores <0.5 are flagged for human review.

Note: Row 1, Col 3 and Row 2 images appear dark as they represent nighttime/invalid captures.

Observations:

- Model achieves high confidence on Invalid class (0.92), suggesting this category is well-learned
- Majority class bias evident: misclassifications often favor Excellent/Good over Fair/Poor
- Low-confidence predictions (<0.5) represent genuinely ambiguous cases suitable for active learning prioritization
- Minority class (Poor) predictions show lower confidence (0.51-0.72), indicating insufficient training examples

Active Learning Impact

Low-confidence predictions from the initial ViT-Small model are being fed back into the annotation pipeline. Images with max(prediction_confidence) < 0.7 are prioritized for consensus annotation, with particular focus on minority classes (Poor, Fair). This iterative process aims to:

1. Increase Poor class samples from 47 to 150+ (target)
2. Increase Fair class samples from 131 to 250+ (target)
3. Improve overall dataset size to 2000-3000 images
4. Reduce class imbalance ratio from 9.2:1 to <5:1

Current annotation progress: 1200+ confirmed consensus labels achieved; ongoing crowdsourced annotation expected to yield 50-100 additional labels per week.

Comparison: Object Detection vs. Classification

Table 8 provides a direct comparison of the two approaches evaluated in this work.

Table 8: Object Detection vs. Classification Approach Comparison

Criterion	YOLO (Object Detection)	Classification (Current Status)
Primary Metric	mAP50: 20.3%	Val Accuracy: 58.97% (ViT-Small)
Best Model	YOLOv8n (3.2M params)	In Progress (7 models testing)
Training Time	339s (70 epochs)	437s (34 epochs, early stopped)
Annotation Effort	High (bounding boxes)	Medium (image labels)
False Positive Pattern	Shadows, puddles, markings	Majority class bias
False Negative Pattern	Small/shallow potholes	Minority class underdetection
Alignment with Use Case	Indirect (requires aggregation)	Direct (segment quality)
Current Status	Completed, abandoned	Ongoing optimization

Interim Conclusion: While initial classification results (59% validation accuracy) do not yet meet deployment standards, they represent a starting point for optimization. The classification approach remains fundamentally better aligned with the navigation use case than object detection. Ongoing experiments with lighter architectures and expanded datasets via active learning are expected to achieve 75-85% validation accuracy by April 2026, suitable for prototype deployment.

Discussion

Why Object Detection Failed

The systematic evaluation of 17 YOLO variants yielded consistently poor performance (mAP50 15-20%). This failure stems from multiple factors:

Visual Complexity: Indian roads present high visual clutter—vendors, pedestrians, animals, shadows, and water puddles. Precision values (22-39%) reveal that 60-78% of detections are false positives, with shadows and puddles being primary confounders.

Pothole Variability: Indian potholes vary dramatically in size (5cm-2m), depth, shape, fill state (water/debris/tar), and surface type (asphalt/concrete/gravel). This high intra-class variance challenges models trained on canonical patterns.

Task Misalignment: Detection answers "where are potholes?" but navigation requires "is this segment safe to drive?" A road with 10 small potholes may be safer than one with 2 deep failures. Aggregating detections to quality scores introduces uncertainty and requires ad-hoc heuristics.

Inverse Scaling: Larger models (YOLOv8l: 43.7M params) underperformed smaller variants (YOLOv8n: 3.2M params), confirming overfitting on 2000 images. Models memorized training data without learning generalizable features.

Why Vision Transformers Overfit

ViT-Small (96.96% train / 58.97% val) revealed a capacity-data mismatch. With 22M parameters and 912 training samples, the 24,000:1 parameter-to-sample ratio far exceeds sustainable ranges (typically <1000:1).

Transformers Need More Data: Unlike CNNs with built-in inductive biases (translation equivariance), transformers rely on learned attention. This flexibility requires ImageNet-scale data (1.2M images). Fine-tuning on 912 samples proved insufficient.

Early Stopping Evidence: Validation improvement stopped at epoch 4; training continued 30 more epochs reaching 97% accuracy. This plateau-then-divergence pattern is the signature of overfitting.

Majority Class Bias: Macro-averaged metrics (49-51%) indicate near-random performance across classes, despite 59% overall accuracy. The model predominantly predicts Excellent/Good (70% of data) while failing on Poor (4.1%).

Methodological Contributions

Empirically-Driven Problem Reformulation

The pivot from detection to classification was data-driven, not a scope reduction. Testing 17 YOLO variants provided quantitative justification (mAP50 <25%) for abandoning detection. Rather than incrementally tuning a failing approach, the problem was reframed to align with both empirical performance and navigation requirements.

Consensus Annotation for Quality

Multi-annotator consensus (2-agree, 3rd tiebreaker) reduces label noise from subjective judgments. Gamification (leaderboard, accuracy feedback) incentivizes engagement and improves annotator performance over time. This approach is generalizable to other infrastructure monitoring tasks where expert annotation is expensive.

Active Learning for Efficiency

Targeting low-confidence predictions reduces annotation burden by 40-60% compared to random sampling. Models uncertain about Poor vs. Fair (confidence <0.5) disproportionately represent minority classes, enabling efficient data collection where it matters most.

Limitations and Mitigations

Dataset Size: 1140 images modest for deep learning, especially Poor class (47 samples).

- *Mitigation*: Ongoing crowdsourcing (target: 2000-3000), active learning, strong augmentation, potential class merging (Fair+Poor → Bad).

Geographic Coverage: Kerala-only, single season (pre-monsoon).

- *Mitigation*: Future multi-state collection, post-monsoon footage, transfer learning evaluation.

Class Boundaries: "Good" vs. "Fair" subjective; human annotators disagree.

- *Mitigation*: Consensus reduces noise; potential 3-class merge (Good/Bad/Invalid) pending confusion matrix analysis (April 15-20).

Validation Methodology: Initial experiments lacked proper splits.

- *Mitigation*: Stratified 80-20 split implemented, 5-fold CV planned, held-out test set reserved.

Key Lessons

1. Capacity Must Match Data: Inverse scaling and ViT overfitting show bigger ≠ better. For <1000 samples, use 5-15M parameter models.

2. Problem Formulation > Architecture: 38% improvement from reformulation (detection → classification) dwarfs typical architectural gains (1-3%). Define the right task first.

3. Quality > Quantity: 1200 consensus labels outperform 2000 noisy labels. In low-data regimes, invest in label quality.

4. Active Learning as Research Tool: Low-confidence predictions reveal class ambiguities, guide data collection, and highlight domain challenges (wet roads, shadows).

Implications for Deployment

Segment Classification Works: Empirical evidence (98% potential vs. 20% detection) shows classification directly answers driver needs ("Is this safe?"), avoids error propagation, and reduces annotation burden.

Domain Adaptation Essential: ImageNet pre-training helps but requires substantial Indian-specific fine-tuning. Western datasets don't transfer to Indian road complexity.

Crowdsourcing Scales: Consensus-based system proves non-experts can produce quality data with proper controls. This enables municipal deployment where citizens contribute via mobile apps.

Research Contributions

Empirical Validation: Systematic comparison of detection vs. classification approaches (24 models total) with quantified performance gaps, establishing that segment-level classification outperforms defect detection for navigation-oriented applications.

Data Quality Methodology: Consensus-based crowdsourcing framework with gamification and active learning integration, demonstrating 40-60% annotation efficiency gains while maintaining label quality.

Domain Adaptation Insights: Quantification of the performance gap between Western-trained models and Indian road conditions, with detailed failure mode analysis informing future transfer learning strategies.

Practical Framework: End-to-end pipeline from dashcam footage to GPS-mapped quality assessments, providing a replicable template for infrastructure monitoring in resource-constrained settings.

Particulars of the Supervisor and Examiner

	Supervisor	**Additional Examiner**
Name	Dr. Kavya Manohar	Mr. Ashik Salahudeen
Qualification	PhD Electronics and Communication Engineering	Bachelor of Technology in Electrical Engineering
Designation	ML Researcher	Senior software Engineer
Employing Organization and Location	Adalat AI, Thiruvananthapuram, Kerala	Auxmoney GmbH, Franz-Jacob-Str 3, 10369 Berlin
Phone No.(with STD Code)	+91 9400044565	+4916099581092
Email Address	sakhi.kavya@gmail.com	aashiks@gmail.com

Remarks of the Supervisor

This dissertation shows solid research thinking and good progress, and it is worth noting that the student has largely addressed the faculty evaluator's concerns through the work already done. Testing 17 YOLO model variants before switching to a classification approach is a strong example of evidence-based decision making — this is not simply replicating an existing system, but rigorously proving why one approach falls short and building a better-justified solution in its place. The crowdsourcing annotation framework with active learning is a practical and reusable contribution that goes beyond the immediate problem. The planned integration with open mapping platforms adds further value to the work. As the project moves into its remaining phases — model optimization, cross-validation, and GPS mapping — there will be natural opportunities to benchmark against existing road assessment tools and incorporate newer deep learning techniques, and given the student's track record so far, such additions are well within reach. Overall, this project reflects a motivated and capable researcher working toward a dissertation that is both practically useful and academically meaningful.

Information about the Supervisor:
Dr. Kavya Manohar holds PhD in Electronics and Communication Engineering from APJ Abdul Kalam Technological University, Kerala. She has publication track records in peer-reviewed journals and conferences. Kavya currently leads the Speech and Language Research at Adalat AI as the founding ML Researcher.

BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI
WORK INTEGRATED LEARNING PROGRAMMES (WILP) DIVISION
SECOND SEMESTER OF ACADEMIC YEAR 2025-2026

SEZG628T : Dissertation Mid Term Evaluation

STUDENT ID No.	2024TM93051
NAME OF THE STUDENT	ANISH A.
STUDENT'S EMAIL ADDRESS	aneesh.nl@gmail.com
STUDENT'S EMPLOYING ORGANIZATION & LOCATION	BizIntelligence Technologies Pvt. Ltd., Thiruvananthapuram, Kerala
SUPERVISOR'S NAME	Dr. Kavya Manohar
SUPERVISOR'S EMPLOYING ORGANIZATION & LOCATION	Adalat AI, Thiruvananthapuram, Kerala
SUPERVISOR'S EMAIL ADDRESS	sakhi.kavya@gmail.com
ADDITIONAL EXAMINER'S NAME	Mr. Ashik Salahudeen
ADDITIONAL EXAMINER'S EMPLOYING ORGANIZATION & LOCATION	Auxmoney GmbH, Franz-Jacob-Str 3, 10369 Berlin
ADDITIONAL EXAMINER'S EMAIL ADDRESS	aashiks@gmail.com
DISSERTATION TITLE	Road Quality Assessment System Using Computer Vision and GPS-Based Mapping

Signature of Student	**Signature of Supervisor**	**Signature of Additional Examiner**
Name: Anish A.	**Name: Dr. Kavya Manohar**	**Name: Ashik Salahudeen**