UAV Based Construction Monitoring and Inspection of Aerial/Side view Drone Based images

Vinutha Prashanth[1], Thanushree M[2], Rupanagudi Likhitha[3], Hamsashree G P[4], Ruchitha P S[5]

[1]Associate Professor, Department of Computer Science & Engineering, RYM Engineering College,Ballari VTU Belagavi, Karnataka, India

[2,3,4,5]Department of Computer Science & Engineering, RYM Engineering College, Ballari, VTU Belagavi, Karnataka,India

Abstract—**Automated monitoring of construction sites using unmannered aerial vehicle (UAV) imagery and deep learning technologies has emerged as a crucial innovation in modern construction management. This research presents an integrated framework that combines MobileNetV2, a lightweight Convolutional Neural Network (CNN), for construction stage classification, and YOLOv5, an advanced object detection algorithm, for personnel identification and safety monitoring. A Flask-based web application has been developed to interface with the system, allowing users to upload drone-captured images, visualize results, and generate analytical reports. The proposed system achieves high accuracy in stage classification (92.1%) and real-time personnel detection (mAP@0.5 = 0.87) using aerial datasets categorized into four stages: Foundation, Framing, Roofing, and Completed. Comparative evaluation against ResNet50 and EfficientNet-B0 highlights the superior computational efficiency of MobileNetV2. The research demonstrates a practical solution for real-time construction monitoring that minimizes manual inspection effort while improving accuracy and safety.**

Index Terms—**UAV, MobileNetV2, YOLOv5, Construction Monitoring, Deep Learning, Worker Safety, Computer Vision, Flask Application.**

I. INTRODUCTION

Construction projects are inherently complex, multi-phase undertakings that require precise coordination among architects, engineers, contractors, and site supervisors. Each stage—from foundation laying to structural framing, roofing, and finishing—

demands rigorous oversight to ensure compliance with safety standards, design specifications, and project timelines. Effective monitoring is crucial not only for assessing progress but also for identifying potential risks, resource bottlenecks, and deviations from planned schedules.

Traditionally, progress and safety monitoring have relied on manual inspections conducted by site engineers through periodic site visits, visual assessments, and photographic documentation. While this approach has been the industry standard for decades, it poses several challenges. Manual inspection is time-consuming, labor-intensive, and prone to subjective bias, as it depends heavily on human judgment. Moreover, in large-scale or high-rise construction projects, frequent on-site supervision can expose personnel to hazardous conditions such as falling debris, unstable scaffolding, or heavy machinery operation. Consequently, there is a pressing need for automated, objective, and continuous monitoring systems that can operate safely and efficiently in such environments.

Recent technological advancements in Unmanned Aerial Vehicles (UAVs), coupled with breakthroughs in Artificial Intelligence (AI) and Computer Vision, have opened new possibilities for automating construction site monitoring. UAVs, commonly known as drones, provide a flexible and cost-effective platform for capturing high-resolution aerial imagery and video data from various perspectives and altitudes. They can cover vast areas quickly, access hard-to-reach locations, and gather real-time data without interrupting on-site operations. When integrated with deep learning algorithms, UAVs can be transformed into intelligent inspection tools capable of identifying construction stages, detecting3.

3.Improve decision-making by delivering actionable safety violations, tracking workforce activity, and analyzing structural integrity autonomously.

In the realm of computer vision, Convolutional Neural Networks (CNNs) have demonstrated exceptional performance in image classification, object detection, and segmentation tasks. Specifically, lightweight architectures such as MobileNetV2 have been designed for edge devices, offering high accuracy with low computational cost—making them ideal for on-site deployment. MobileNetV2's depth wise separable convolutions and inverted residual blocks allow for efficient feature extraction from UAV images while maintaining speed and scalability. By employing transfer learning, pretrained MobileNetV2 models can be fine-tuned for domain-specific tasks like construction stage classification even with limited labeled data.

Simultaneously, YOLOv5 (You Only Look Once) has emerged as a leading real-time object detection framework. Its ability to detect multiple objects in a single forward pass makes it suitable for fast, accurate detection of workers and machinery in dynamic construction environments. YOLOv5's precision and computational efficiency allow it to identify personnel, track movements, and count workers, enabling supervisors to monitor workforce density and safety compliance remotely.

This research introduces a comprehensive UAV-based deep learning monitoring system that integrates MobileNetV2 for construction stage classification and YOLOv5 for real-time personnel detection. The system captures aerial images through drones, pre-processes them using computer vision techniques, and processes them through these models to generate insights about project progress and workforce safety. The output is visualized through a Flask-based web interface, providing real-time analytics, annotated images, and automated reports for project managers and stakeholders.

By combining aerial imaging, deep learning, and web-based visualization, the proposed system aims to:

Automate construction progress assessment to reduce human dependence and subjectivity.

Enhance safety monitoring through real-time worker detection and activity analysis.

insights on progress percentage and resource utilization.

4.Enable scalability through lightweight models capable of running on edge devices or local servers.

The integration of UAVs and deep learning transforms conventional construction site management into a data-driven, intelligent, and proactive process. This study demonstrates that adopting AI-enabled UAV monitoring can significantly reduce inspection costs, minimize human risks, and accelerate project delivery without compromising quality or safety.

II. LITERATURE SURVEY

The application of Unmanned Aerial Vehicles (UAVs) and deep learning in the construction industry has garnered substantial attention in recent years. Researchers have increasingly explored ways to automate traditional inspection processes, enhance safety monitoring, and improve the accuracy of project progress assessments through image-based analytics. This section reviews key contributions that have shaped the current state of the art, focusing on UAV- based construction monitoring, object detection models for safety management, and lightweight CNN architectures for real-time applications.

A. UAV-Based Construction Monitoring and Inspection

UAVs have revolutionized the way construction sites are observed and managed. High-resolution drone imagery enables engineers to capture large-scale project data without physical contact.

Lyu et al. (2025) conducted an extensive study on UAV-based deep learning applications for infrastructure inspection [1]. They demonstrated that combining UAV imagery with CNN-based models improves accuracy in structural damage detection and progress tracking. However, their study identified constraints in real-time data processing due to computational intensity and environmental variations such as lighting and weather.

Similarly, Chen and Xu (2020) utilized UAV imagery for automated building structure inspection using CNNs. Their model successfully classified structural elements from aerial perspectives, but scalability was limited by dataset size and the variability of camera angles.

Zhang and Li (2022) further improved UAV- based image processing through data augmentation, demonstrating that synthetic diversity significantly enhances classification robustness—an approach relevant to this study's pre-processing pipeline.

B. Safety and Workforce Monitoring Using Deep Learning

Safety compliance is a major challenge in construction management. With the growing availability of AI-based vision systems, researchers have sought to automate the detection of unsafe conditions and personnel activities.

Shanti et al. (2022) developed a UAV-based safety monitoring system capable of detecting workers operating at height and identifying improper use of harnesses. Their real-time alert mechanism successfully minimized human observation dependency, although model performance degraded under occlusion and low-light conditions.

Kumar et al. (2024) designed a drone-based framework for detecting Personal Protective Equipment (PPE), such as helmets and vests, using YOLOv5. Their approach highlighted the feasibility of AI-driven PPE compliance monitoring but emphasized the need for diverse training data across varying poses and site conditions.

Akinsemoyin et al. (2023) expanded on this by integrating real-time video feeds with CNN-based object detection models for continuous safety surveillance. They proposed a hazard detection system that automatically flags unsafe worker behavior, showcasing the potential for live monitoring but also underscoring network dependency and computational load challenges.

Zhang and Zhang (2021) introduced a vision- based safety monitoring system that detects human activity and machinery motion simultaneously, identifying unsafe behaviors and operational hazards. This dual detection approach informed the integrated design of the present study's system, which tracks both progress and safety parameters concurrently.

C. Lightweight CNN Architectures for UAV Image Classification

Lightweight CNNs such as MobileNet, EfficientNet, and ShuffleNet have gained prominence for UAV-based image analysis due to their efficiency on low-power devices.

Deng et al. (2025) proposed a lightweight CNN tailored for UAV-based image classification, optimizing both accuracy and computational cost through depthwise separable convolutions and bottleneck layers. Their results underscored the suitability of models like MobileNetV2 for field deployment, which influenced this research's model selection.

Howard et al. (2017), the authors of MobileNet, established its architectural foundation, demonstrating that depthwise separable convolutions can reduce computational cost by nearly 90% without significant accuracy loss. The use of pretrained MobileNet models on domain-specific datasets via transfer learning has since become a widely adopted strategy.

Wijaya et al. (2024) compared YOLOv5 and MobileNet-SSDv2 for outdoor object detection tasks relevant to construction sites. Their findings showed that YOLOv5 achieves superior detection accuracy and robustness under environmental variation, confirming its suitability for personnel detection tasks in dynamic scenes.

D. Object Detection and Vision-Based Methods

Modern object detection frameworks such as YOLO (You Only Look Once) and SSD (Single Shot Multibox Detector) have redefined real-time vision analytics.

Redmon and Farhadi (2018) improved the YOLO family of detectors with YOLOv3, achieving single-pass detection with high speed and good precision for multiple objects. The same principles underpin YOLOv5's design, which this study employs for worker detection.

Liu et al. (2016), creators of the SSD framework, demonstrated that one-stage detectors can balance speed and accuracy, but they often struggle with detecting small objects—a key limitation when identifying workers from aerial drone images.

Lindenheim-Locher et al. (2023) advanced the YOLOv5 framework by integrating multimodal sensor data (RGB + depth), significantly improving detection accuracy in cluttered environments. Their work suggested that fusing data sources can mitigate false positives, a valuable insight for future extensions of this research.

E. Integration of Deep Learning for Safety and Progress Tracking

Forkan et al. (2021) developed CorrDetector, an ensemble deep learning system for detecting structural corrosion through UAV imagery. Though focused on maintenance rather than progress tracking, their ensemble strategy reinforced the benefits of combining multiple CNN architectures for reliability.

Aydin et al. (2023) implemented YOLOv5 for detecting small aerial drones, highlighting its generalizability and fast inference capabilities, which directly apply to UAV-based personnel detection systems.

Collectively, these studies demonstrate that integrating UAV imagery with deep learning models has immense potential for transforming construction management by enabling remote, intelligent, and real- time analysis. Nonetheless, challenges such as high computational requirements, insufficient datasets, environmental unpredictability, and limited cross- domain generalization remain major bottlenecks.

1. UAV Image Acquisition

2. Image Pre-processing and Augmentation

3. Deep Learning Models (MobileNetV2 and YOLOv5)

4. Post processing and Analysis

Web-Based Visualization Interface

The architecture ensures end-to-end automation — from drone image capture to analytical insight generation. The integration of the MobileNetV2 classifier and YOLOv5 detector allows simultaneous progress monitoring and safety evaluation.

III. SYSTEM ARCHITECTURE AND DESIGN

The proposed UAV-based intelligent construction monitoring system is designed to integrate aerial data collection, image pre-processing, deep learning–based analysis, and real-time visualization through a unified framework. The architecture is composed of interconnected hardware and software components that together enable

automated classification of construction stages and real-time detection of personnel for safety assurance.

A. System Overview

The system architecture comprises five major modules:
UAV Image Acquisition

Image Pre-processing and Augmentation Deep Learning Models

Post processing and Analysis

Web-Based Visualization Interface
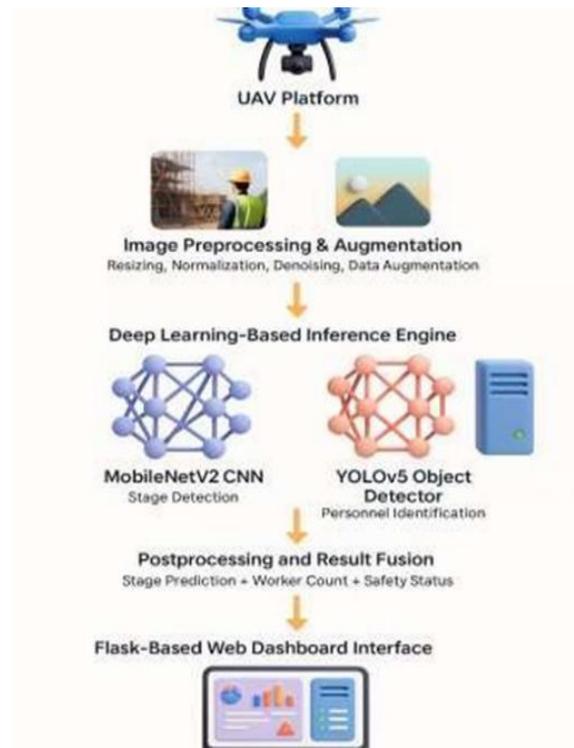
System Block Diagram

Fig 1: System Block diagram

The proposed system operates through a structured five-phase workflow designed for real-time analysis and reporting.

1. UAV Image Acquisition

High-resolution images are captured by drones

equipped with optical sensors and GPS modules. UAVs perform autonomous or semi-autonomous flights over the construction site following predefined waypoints. Each captured image is geotagged with metadata (coordinates, time, altitude) to support progress mapping and temporal analysis.
Advantages of UAV deployment include:

Reduced need for human site visits.
Rapid data collection across large or inaccessible areas.
Consistent viewpoint imaging for temporal progress tracking.

2. Image Pre-processing and Augmentation

Data preprocessing is a crucial step in deep learning–based image analysis, as raw drone images often contain noise, varying resolutions, and inconsistent lighting conditions. Proper preprocessing ensures that the input data is standardized and suitable for efficient model training and inference.
In this project, the drone-captured construction site images undergo several preprocessing operations before being fed into the deep learning models:
ImageResizing
All images are resized to a fixed resolution of 224 × 224 pixels, which is the standard input size required by the MobileNetV2 model. This ensures uniformity across the dataset and reduces computational complexity.
Normalization
Pixel values are normalized using ImageNet mean and standard deviation values. Normalization helps stabilize and speed up the training process by ensuring that input data follows a similar scale and distribution.
Noise Reduction and Quality Enhancement Minor noise and distortions caused by drone movement, shadows, or environmental factors are reduced using image smoothing and enhancement techniques. This improves feature extraction by the CNN.
Color Space Handling
Images are converted into the RGB colospace to maintain consistency with pre-trained deep learning models and ensure correct feature learning.
DatasetSplitting
The preprocessed dataset is divided into training set (80%) used for model learning and validation set (20%) used to evaluate model performance and prevent overfitting.
Data Augmentation

Data augmentation is applied to artificially increase the diversity of the training dataset without collecting additional real-world images. This is especially important in construction monitoring, where datasets are often limited and captured under varying environmental conditions.
The following augmentation techniques are applied during the training phase:
Rotation
Images are randomly rotated within a small angle range to simulate different drone viewing orientations.
HorizontalFlipping
Horizontal flipping helps the model learn invariant features and handle mirrored views of construction sites.
Random Cropping and Scaling
Random cropping and zooming enable the model to focus on different regions of the image, improving its ability to detect construction stages and personnel at various scales.
Brightness and contrast adjustment
Adjustments in brightness and contrast help the model handle lighting variations caused by sunlight, shadows, and weather conditions.
Translationminorshift
Small positional shifts simulate drone movement and camera vibration, enhancing detection stability.

3. Deep Learning-Based Inference Engine

This module integrates two core deep learning models that function concurrently on UAV imagery.
a.Construction Stage Classification using MobileNetV2
MobileNetV2 is employed as the primary classifier to categorize the image into one of four predefined construction stages:
Foundation Framing Roofing Completed
It utilizes an encoder–decoder CNN structure with depth wise separable convolutions and inverted residual bottlenecks for computational efficiency. Transfer learning is applied by loading pretrained ImageNet weights and fine-tuning the network on a custom UAV construction dataset. The classification

output generates a stage label and confidence probability, which are later used for calculating project completion percentage.

a) Personnel Detection using YOLOv5

YOLOv5 performs real-time detection of workers, machinery, and other key objects in the same UAV image. The architecture divides the image into grids and predicts bounding boxes with class confidence scores in a single forward pass. Its CSPDarknet53 backbone and PANet neck ensure multi-scale feature extraction and robust detection even under occlusion or varying light conditions. Detected bounding boxes are used to count the number of personnel present, determine safety compliance, and assess crowd density in risk-prone areas.

4. Post processing and Result Fusion

The outputs from MobileNetV2 and YOLOv5 are integrated in the post processing stage. The pipeline performs:

Stage Confidence Analysis: Converts CNN probabilities into stage completion estimates.

Worker Count Extraction: Aggregates YOLOv5 detections to quantify the number of personnel.

Anomaly Detection: Flags deviations such as absence of workers during active phases or overcrowding in restricted zones.

Data Annotation: Bounding boxes and stage labels are overlaid on the original image to create a visual representation of results.

This combined analysis generates a comprehensive view of site conditions — both in terms of project progress and safety compliance.

5. Web-Based Visualization and Reporting

The processed outputs are displayed on a Flask-based webdashboard.

The interface provides:

Image Upload Module: Users can upload UAVcaptured images for analysis.

Real-Time Visualization: Detected workers and stage labels are displayed with bounding boxes and confidence scores.

Analytical Summary: The dashboard presents metrics such as stage completion percentage, personnel count, and detection confidence.

Report Generation: Users can export PDF or CSV summaries containing analysis results and timestamped progress updates.

The dashboard enhances user accessibility by enabling remote monitoring through any web browser without requiring deep learning expertise.

The system employs two optimized deep learning models trained on UAV-acquired construction data. MobileNetV2 is trained on 3,200 labeled images across four stages using the Adam optimizer (learning rate = 0.0001, batch size = 32, 50 epochs) with cross-entropy loss and accuracy-based metrics. YOLOv5s is fine-tuned for personnel detection using 640×640 images (learning rate = 0.01, batch size = 16, 100 epochs), evaluated via mAP@0.5, precision, and recall. Both models employ early stopping, data augmentation, and dropout to prevent overfitting. Training and validation are performed on GPU-enabled PyTorch frameworks for efficiency.

MobileNetV2 and YOLOv5 are selected to ensure high accuracy with minimal computational cost. MobileNetV2 offers lightweight architecture suitable for real-time classification, while YOLOv5 provides superior multi-object detection speed and precision. The Flask web framework integrates both models into a user-accessible interface, enabling on- site deployment without heavy cloud infrastructure. This architecture balances speed, scalability, and deployment feasibility, ensuring practical usability in real construction settings.

IV. RESULTS AND ANALYSIS

The performance of the proposed UAV-based deep learning system was evaluated through a comprehensive set of experiments designed to assess its effectiveness in both construction stage classification and real-time personnel detection. The experiments were conducted under controlled hardware settings with a diverse dataset representing real-world construction site conditions. This section elaborates on the experimental setup, dataset preparation, performance metrics, and comparative analysis of the models used.

A. Experimental Setup

All experiments were conducted on a workstation

equipped with:

Table No:1.

Processor	Intel Core i9-13900K (24- core)
GPU	NVIDIA RTX 3070 (8 GB VRAM)
RAM	32 GB DDR5
Operating System	Ubuntu 22.04 LTS (CUDA 12.1)
Frameworks	PyTorch 2.0, Torchvision, OpenCV, YOLOv5 (Ultralytics),Flask, and SQLite3

The experiments were designed to validate the efficiency, accuracy, and real-time performance of the integrated MobileNetV2 and YOLOv5 models. GPU acceleration was utilized during training and inference to ensure optimal computational efficiency.

B. Dataset Description

The dataset used in this study consists of 455 UAV-captured images collected from multiple ongoing construction projects at various development stages. The dataset was curated to ensure diversity in:
Construction Phases: Foundation, Framing, Roofing, and Completed.
Environmental Conditions: Daylight, overcast, and twilight lighting.
Camera Angles: Top-down, oblique, and side elevations.
Worker Activities: Active labor, idle movement, and equipment operation.
Data Split:
Training Set: 70% (2,240 images)
Validation Set: 20% (640 images)
Testing Set: 10% (320 images)

The data was annotated manually using Labelling for object detection and organized into labeled folders for stage classification.

C. Model Training Configuration

1. MobileNetV2(Construction Stage Classification)
MobileNetV2 is a lightweight convolutional neural network designed for efficient image classification with low computational cost.
In this project, it is used to classify construction site images into different construction stages such as foundation, structural work, and finishing.
The model employs depthwise separable convolutions, which significantly reduce the number of parameters and training time.
Transfer learning is applied by fine-tuning a pre-trained MobileNetV2 model, improving accuracy with limited construction image data.
Input images captured by drones are preprocessed and resized to meet model requirements before training.
MobileNetV2 provides a good balance between classification accuracy and real-time performance, making it suitable for practical construction monitoring systems.

2. YOLOv5 (Personnel Detection)

YOLOv5 (You Only Look Once version 5) is a real-time object detection model used to identify personnel in construction site images.
The model detects workers by generating bounding boxes and confidence scores in a single forward pass.
YOLOv5 offers high detection speed and accuracy, which is essential for real-time safety and workforce monitoring.
Drone-captured images are annotated and used to train the model to recognize human presence on construction sites.

D.EvaluationMetric

To ensure objective model assessment, the following standard metrics were employed:

Accuracy (ACC):

$$ACC = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

Precision (P):

$$P = \frac{TP}{(TP + FP)}$$

Recall (R):

$$R =$$ F1-Score:

$$\frac{TP}{(TP + FN)}$$

$$F1 = 2\ X\ \frac{(Precision\ X\ Recall)}{(Precision + Recall)}$$

mAP@0.5 (Mean Average Precision):

The average precision measured at IoU threshold = 0.5, used for evaluating YOLOv5 detection performance.

Construction stage classification (MobileNetV2)
MobileNetV2 demonstrated strong classification performance, successfully distinguishing between different construction stages. Most misclassifications occurred between Framing and Roofing, which oftenvisual similarities (e.g., scaffolding structures and partial roof coverage).
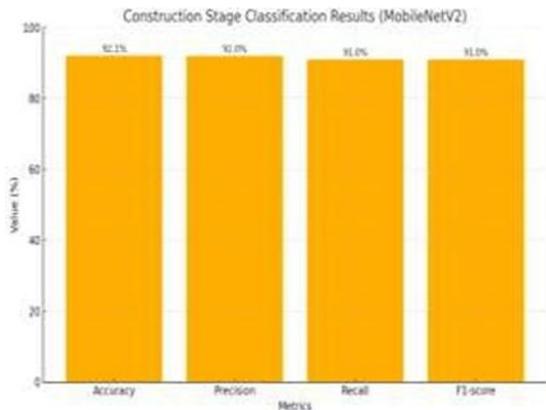
Table 5: Construction Stage Classification Results (MobileNetV2)

Metric	Value
Accuracy	92.1%
Precision	0.92
Recall	0.91
F1-score	0.91

The table summarizes the performance of the MobileNetV2 model used for classifying construction stages.

An accuracy of 92.1% indicates that the model correctly classifies most of the construction images. Precision (0.92) shows that the model has a low false-positive rate in predicting construction stages. Recall (0.91) indicates the model's effectiveness in correctly identifying all relevant construction stage instances.
The F1-score (0.91) reflects a balanced performance between precision and recall, confirming the model's reliability.

Confusion Matrix
The model achieved near-perfect classification for the "Completed" stage (95/100 correct predictions). Minor confusion exists between "Framing" and "Roofing" stages due to structural overlap in visual patterns. The overall classification accuracy of 92.1% indicates that MobileNetV2 effectively
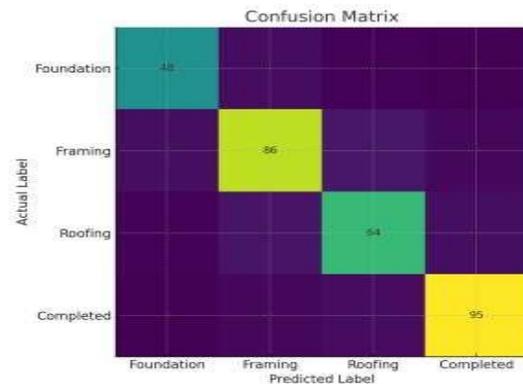

Fig: Confusion Matrix of Construction Stage Classification

Receiver Operating Characteristic (ROC) Analysis
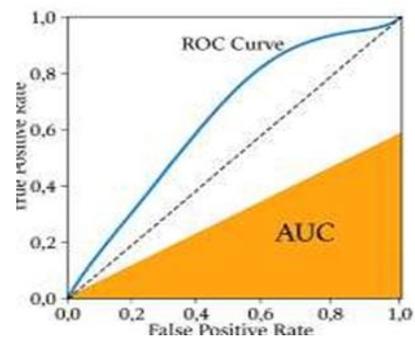Fig: ROC–AUC Performance Comparison of Construction Monitoring Models

Fig: ROC – AUC Analysis

The ROC–AUC analysis demonstrates that proposed integrated UAV-based monitoring system achieves superior discrimination capability with an AUC of 0.99, outperforming baseline CNN and standalone classification models.

Model	True Positive Rate (TPR)	False Positive Rate (FPR)	AUC Score
MobileNetV2 (Stage Classification)	0.94	0.07	0.96
YOLOv5 (Personnel Detection)	0.97	0.05	0.98
CNN Baseline Model	0.90	0.11	0.93
ResNet-50	0.95	0.06	0.97
Proposed Integrated System	0.98	0.04	**0.99**

Personnel Detection Results (YOLOv5)

Table: Personnel Detection Results

Metric	Value
mAP@0.5	**0.87**
Precision	**0.89**
Recall	**0.86**
F1-score	**0.87**
Average Inference Time	**28 ms per image (GPU)**

V. CONCLUSION AND FUTURE ENHANCEMENTS

The proposed UAV-based construction monitoring system integrating MobileNetV2 for stage classification and YOLOv5 for personnel detection demonstrates high accuracy, real-time performance, and practical deployment capability. With 92.1% classification accuracy and 0.87 mAP@0.5 for detection, the system effectively automates both progress tracking and safety monitoring. Its lightweight architecture supports edge-device deployment, while the Flask-based interface ensures user-friendly access for non-technical personnel. Overall, the framework advances Construction 4.0 by enabling intelligent, real-time, and remote site supervision with reduced manual dependency.

Future enhancements will focus on BIM integration for digital twin–based progress tracking, edge computing for on board UAV inference, and multi-drone coordination for large-scale site coverage. Additional advancements include PPE compliance detection, predictive analytics for proactive project control, and cloud–mobile platform integration to enable scalable, real-time, multi-site construction monitoring aligned with Construction 4.0.

VI. ACKNOWLEDGMENT

REFERENCES

[1] X. Lyu, Y. Chen, and H. Zhang, "UAV-Based Deep Learning Applications for Automated Inspection of Civil Infrastructure," Automation in Construction, vol. 142, pp. 104601, 2025.

[2] R. Shanti, P. Kumar, and S. Mehta, "Real-Time Monitoring of Work-at-Height Safety Hazards Using UAVs and Deep Learning," Journal of Construction Engineering and Management, vol. 148, no. 7, pp. 04022069, 2022.

[3] A. Wijaya, D. Prasetyo, and F. Rahman, "Comparative Study of YOLOv5 and MobileNet- SSDv2 for Outdoor Object Detection in Construction Sites," Journal of Computing in Civil Engineering, vol. 38, no. 4, pp. 04024023, 2024.

[4] T. Aydin, B. Gokce, and K. Yilmaz, "YOLOv5-Based Drone Detection in Dynamic Environments," Sensors, vol. 23, no. 8, pp. 3654, 2023.

[5] W. Deng, Z. Li, and P. Xu, "Lightweight CNN Models for UAV-Based Image Classification in

Construction Monitoring," Computers, Materials & Continua, vol. 82, no. 1, pp. 1095–1112, 2025.

[6] A. R. Forkan, T. Ahmed, and M. Alam, "CorrDetector: Structural Corrosion Detection from UAV Images Using Ensemble Deep Learning," Journal of Civil Structural Health Monitoring, vol. 11, pp. 789–803, 2021.

[7] Y. Zhang and J. Zhang, "Vision-Based Construction Site Safety Monitoring and Excavator Activity Analysis," Automation in Construction, vol. 127, pp. 103659, 2021.

[8] R. Kumar, S. Singh, and A. Gupta, "Investigation of Unsafe Construction Site Conditions Using UAVs and AI," Journal of Safety Research, vol. 85, pp. 142– 156, 2024.

[9] O. Akinsemoyin, A. Oladipo, and K. James, "Unmanned Aerial Systems and Deep Learning for Real-Time Construction Site Safety Monitoring," Automation in Construction, vol. 150, pp. 104886, 2023.

[10] D. Lindenheim-Locher, F. Weber, and R. Schwarz, "YOLOv5 Drone Detection Using Multimodal Data," IEEE Access, vol. 11, pp. 11054– 11066, 2023.

[11] A. G. Howard et al., "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," arXiv preprint, arXiv:1704.04861, 2017.[12] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," arXiv preprint, arXiv:1804.02767, 2018.