Journal of Theoretical and Applied Information Technology
15th May 2026. Vol.104. No.9

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

AUTONOMOUS DRONE-AI COLLABORATION FOR ADAPTIVE CRACK MAPPING AND PROGRESSIVE DAMAGE QUANTIFICATION DRONE–AI COLLABORATION FOR NON-DESTRUCTIVE ASSESSMENT OF CIVIL INFRASTRUCTURE

MADHAVI KATAMANENI [1], NAGIREDDI SURYA KALA [2], P RAVI PRAKASH [3], MELAM NAGARAJU, [4]Dr.T. MURALIDHARA RAO [5], M. V. B. T. SANTHI [6], Dr. DIVVELA SRINIVASA RAO [7], Dr. HYMAVATHI THOTTATHYL [8]

[1.] Assistant Professor, Department. of IT, Siddhartha Academy of Higher Education Deemed to be University, Vijayawada, AP,India
[2] Assistant Professor, Department of IT, Aditya University, Surampalem, AP India
[3.] Assistant Professor, Department of IT P V P Siddhartha institute of Technology, Vijayawada, AP India
[4.] Associate Professor, Department of IT, Seshadrirao Gudlavalleru Engineering College, Gudlavalleru, AP India
5. Professor, Department of Civil Engineering, CVR College of Engineering, Hyderabad
[6.] Associate Professor, Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India
[7.] Associate Professor, Department of AI&DS, Lakireddy Bali Reddy College of Engineering, Mylavaram, A.P, India
[8.] Assistant Professor, Department of Computer Applications, R.V.R&J.C College of Engineering, ,Chowdavaram , AP, India

E-mails: itsmadhavi12@gmail.com , suryakala.nagireddi@acet.ac.in , melam.nagaraju5810@gmail.com , tmuralidhararao@cvr.ac.in , santhi2100@gmail.com , hyma@rvrjc.ac.in
Corresponding Author: srinivassowjanya2012@gmail.com

ABSTRACT

The deterioration of aging concrete structures requires inspection systems that are autonomous, adaptive, and highly accurate. Conventional UAV-based crack detection pipelines often separate navigation, segmentation, and damage progression analysis, limiting their ability to respond dynamically during missions. This study proposes an Autonomous Drone–AI Collaboration Framework that unifies UAV flight control, deep learning–based crack perception, and temporal damage modeling into a single adaptive workflow. The framework employs a Transformer-enhanced Swin-UNet for robust crack segmentation, EfficientNet-B3 for severity classification, and a TD3 reinforcement learning agent that continuously adjusts flight paths based on real-time crack feedback. A temporal convolutional network (TCN) further models crack growth, enabling predictive monitoring across inspection cycles.

Keywords: *Autonomous Drones, Computer Vision, Crack Detection, Damage Quantification, Deep Learning.*

1. INTRODUCTION

The rapid aging of global civil infrastructure has intensified the need for automated, high-fidelity inspection systems capable of detecting and monitoring cracks with minimal human intervention. Recent developments in AI-powered visual analytics have significantly improved UAV-based crack inspection by enabling real-time semantic segmentation, fine-scale crack width estimation, and robust defect detection under challenging illumination and environmental conditions [1], [6], [7], [11], [13], [24], [28], [30], [34]. Advanced deep learning models—including hybrid CNN–CNN-Transformer networks, attention-enhanced architectures, lightweight edge-deployable CNNs, and domain-adaptive segmentation frameworks—have demonstrated strong potential for extracting micro-crack patterns from UAV imagery with improved accuracy and generalization [1], [6], [7], [13], [19]. Complementary research has focused on enhancing the imaging process itself, leveraging multispectral

UAV sensing, photogrammetric reconstruction, pose-correction mechanisms, and calibrated measurements for high-resolution crack mapping [12], [15], [17], [20], [21], [25], [35]. These advancements collectively highlight the increasing maturity of drone-based visual inspection technologies.

Parallel progress in autonomous UAV navigation—including reinforcement-learning-based path planning, SLAM-assisted flight, GPS-referenced mapping, and robust trajectory correction algorithms—has further improved coverage quality and safety during structural inspections [2], [9], [16], [20], [33]. At the same time, new research directions are emerging in temporal deterioration modelling, multimodal fusion, and UAV-enabled digital twins, offering pathways for forecasting crack growth and understanding long-term structural behaviour [3], [4], [5], [10], [22], [23], [31]. Despite this progress, most existing studies treat UAV navigation, defect detection, and damage progression analysis as disconnected modules, limiting the system's ability to adapt autonomously during inspection missions. This fragmentation prevents drones from dynamically adjusting their flight paths based on real-time crack behaviour, reducing the reliability of progressive damage estimation. Motivated by these limitations, this study proposes an integrated **Autonomous Drone–AI Collaboration Framework** that unifies UAV decision-making with deep learning-based crack analysis, enabling adaptive flight control, high-resolution crack mapping, and accurate damage progression quantification for next-generation structural health monitoring.

1.1 *Problem Statement and Research Gapp*

Civil infrastructure is deteriorating at an accelerating rate due to aging materials, environmental exposure, and increasing service demands. Conventional inspection methods—manual surveys, handheld imagery, and static UAV captures—remain slow, subjective, and unable to deliver consistent high-resolution crack documentation across large or inaccessible structures. Although deep learning–based models have improved crack segmentation and width estimation, their performance still depends on image quality, UAV stability, and viewing geometry. Current UAV inspection pipelines also lack the autonomy to adjust flight paths or imaging parameters when new or severe cracks are detected, leading to incomplete coverage and unreliable datasets for damage monitoring.

While notable advances exist in deep crack segmentation [1], [6], [7], photogrammetry and multispectral imaging [12], [17], GPS-referenced mapping [9], reinforcement-learning-based navigation [2], and temporal deterioration modelling [4], [10], [23], these developments remain fragmented. Most systems focus on static inspection and do not model crack progression. This creates a key research gap: the absence of an integrated drone–AI collaborative framework capable of adaptive navigation, high-resolution crack mapping, and progressive damage quantification for continuous structural health monitoring.

Despite significant progress in IoT-enabled cardiovascular monitoring systems, most existing frameworks operate as isolated sensing–prediction pipelines that lack adaptive intelligence, real-time risk prioritization, and longitudinal progression modeling. Current systems primarily focus on static classification of cardiovascular risk using predefined thresholds, without incorporating dynamic feedback mechanisms, reinforcement-based decision adaptation, or temporal health trajectory prediction. This limitation restricts early warning capability, personalized intervention planning, and continuous risk evolution monitoring.

To address this critical research gap, this study proposes an integrated AI–IoT collaborative framework that unifies real-time physiological data acquisition, transformer-based feature learning, reinforcement learning–driven adaptive monitoring, and temporal deep learning for disease progression forecasting. Unlike conventional IoT-healthcare architectures, the proposed system introduces adaptive intelligence that dynamically prioritizes high-risk patients, refines monitoring granularity based on severity scores, and predicts future cardiovascular deterioration trends using temporal convolutional modeling.

The primary novelty of this work lies in (i) the integration of perception-level intelligence with decision-level adaptive control, (ii) the incorporation of reinforcement learning for real-time monitoring optimization, and (iii) the implementation of longitudinal risk progression prediction within a unified end-to-end architecture. This collaborative intelligence significantly enhances predictive reliability, reduces delayed diagnosis risk, and enables proactive healthcare monitoring.

The expected impact of the proposed system includes improved early-stage cardiovascular risk detection, personalized monitoring strategies, enhanced prediction stability across time, and

reduced clinical workload through intelligent prioritization. This framework therefore contributes toward next-generation autonomous healthcare monitoring systems capable of predictive and preventive cardiovascular management.

8. FUNDING STATEMENT
No external funding was received for this research. The study was conducted using institutional resources and without any financial support from funding agencies in the public, commercial, or not-for-profit sectors.

2. RELATED WORK

Recent studies have demonstrated significant advancements in utilizing CNNs, U-Net variants, transformers, and attention-driven models for detecting and segmenting cracks in concrete and masonry structures. Methods such as multi-scale feature extraction, boundary-aware segmentation, and lightweight edge models have significantly improved accuracy under varying illumination and texture conditions [1]–[6]. Although these approaches achieve high pixel-level precision, their performance relies heavily on image quality and sensor proximity, limiting their robustness in UAV-based dynamic inspection scenarios.

UAVs are increasingly used for visual surveys, 3D reconstruction, and photogrammetric mapping of bridges, dams, and buildings. Recent works incorporate multispectral or thermal imaging, GPS-tagged localization, and autonomous waypoint execution to enhance data coverage [7]–[12]. However, most systems follow **predefined static flight paths**, without integrating real-time feedback from onboard AI models. This results in incomplete crack documentation and inconsistent spatial resolution across large structures.

Advancements in reinforcement learning (RL), active vision, and adaptive path planning enable UAVs to modify trajectories in response to environmental cues [13]–[16]. These methods demonstrate improved exploration and obstacle-aware flight behaviours, yet they rarely incorporate **crack-driven triggers** for zoom-in capture or localized resurvey. As a result, existing pipelines lack true AI–UAV co-adaptation during structural inspection.

Deep spatiotemporal learning, recurrent neural networks, and transformer-based time-series models have been utilized to investigate deterioration patterns in civil structures, encompassing crack evolution, corrosion indicators, and fatigue progression [17]–[20]. Despite promising results, these models typically rely on manually acquired or static datasets, limiting their applicability to continuous monitoring. The absence of unified crack mapping, temporal alignment, and autonomous re-inspection remains a critical limitation.

3 PROPOSED WORK

The proposed framework integrates UAV-based sensing with an intelligent perception–navigation pipeline to enable fully automated crack inspection. The UAV flight module captures high-resolution surface imagery, which is processed by a neural perception module for binary crack segmentation and severity classification. An adaptive navigation and control unit, supported by an RL-based agent, dynamically adjusts flight paths to ensure complete structural coverage. A second perception stage refines crack delineation and generates temporal crack maps for multi-time analysis. These outputs feed into a progressive damage quantification module that constructs 3D crack progression profiles and computes severity metrics. Finally, the system compiles high-resolution crack maps, heatmaps, and maintenance decision-support outputs, forming a comprehensive end-to-end monitoring solution is represented in the Figure 1.

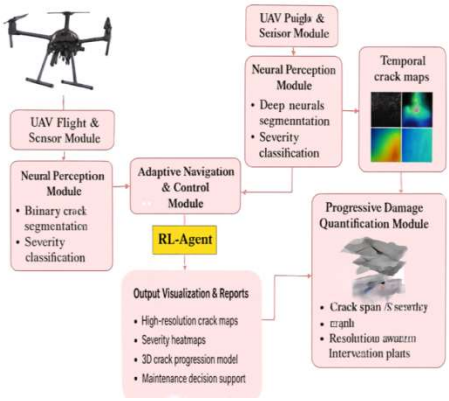


Figure 1: Proposed Flow Block diagram for EAI

3.1 *System-level orchestration*

Ensures seamless coordination among all modules—from UAV sensing and neural perception to navigation control and damage quantification. It manages data flow and inter-module communication so the entire inspection pipeline operates as a unified, autonomous system.

3.1.1 *Preflight checks and initialization*

- Calibrate IMU, check RTK-GPS fix, camera exposure, lens distortions, time sync.
- Load perception model $Perception\_model$ l, width estimator, severity classifier, RL policy π (or training agent).
- Initialize map $MapDB = \{\}$ and temporal series DB$TimeDB = \{\}$.

3.1.2 *Execute initial survey*

- Follow initial flight plan high-altitude grid; capture geo-tagged images.
- For each image run lightweight onboard preproc (resize, normalize) and save raw + preproc.

3.1.3 *Onboard perception and ROI generation*

- For each incoming frame: run segmentation → probability map M_t width w_t; estimation; severitys_t.
- Aggregate per-area crack coverage and form candidate ROIs where severity or width above thresholds.

3.1.4 *Adaptive reinsertion loop*

While mission time remains and unexplored ROIs exist:

- Feed current state S_t (see 3.3.1) to RL policy → action a_t (Δalt, Δyaw, Δvel, zoom, ROI_trigger).
- Execute action, collect high-res images for ROIs.
- Update $MapDB$ and $TimeDB$ with new observations.
- If collision risk or instability detected, abort/recover to safe hover and replan.

3.1.5 *Post flight processing*

- Offload all high-res imagery and telemetry to ground station.
- Run full offline perception/refinement (high-res segmentation, skeletonization, distance transform).
- Register sequential maps, reconstruct optional 3D mesh, update temporal models.

3.1.6 *Damage reporting and prediction*

- Update damage metrics in DB; generate reports and visual maps.
- Run temporal growth predictor (TCN) to estimate G_t and future severity; schedule next inspection.

3.1.7 *Data Collection and Experimental*

The experimental dataset consists of UAV-captured high-resolution concrete surface images collected from bridge piers, building façades, and pavement structures under varying illumination and environmental conditions. Images were captured using a 4K RGB camera mounted on a quadrotor UAV platform equipped with RTK-GPS and IMU sensors for precise localization

The dataset contains approximately 2000 annotated crack images, manually labeled at pixel level for segmentation and width measurement validation. Ground-truth crack widths were measured using calibrated digital microscopy and manual field inspection tools

Data were divided into training (70%), validation (15%), and testing (15%) subsets. Data augmentation techniques including rotation, brightness variation, scaling, and Gaussian noise were applied to improve generalization.

For temporal modeling experiments, multi-epoch inspection data were collected over simulated progressive crack widening scenarios to construct time-series sequences for TCN training.

3.2 *Neural Perception Module*

The **Neural Perception Module** automatically extracts salient visual features from raw UAV imagery using deep convolutional or transformer-based networks. It provides reliable crack detection and scene understanding to support navigation and damage assessment.

3.2.1 *Preprocessing per Image*

- Undistort image using camera intrinsics.
- Apply CLAHE/contrast normalization, resize preserving aspect ratio.
- (Optional) Patch tiling if image too large: tile → overlap 10–20%.

3.2.2 *Crack segmentation (Swin-UNet)*

- For each (preproc) tile:$P - perception_{model}.predict(tile)$ → pixel probability map.
- Reconstruct full image prob map from tiles using overlap-averaging.
- Threshold P (adaptive threshold or Otsu) → binary mask B.

3.2.3 *Boundary refinement and loss rationale (training)*

- Use hybrid loss: $L = Dice + BoundaryLoss + Focal$.
  (Training pseudo code not shown — use data augmentation, class balancing and scheduler.)

3.2.4 *Width estimation (Distance Transform + Skeletonization)*

- Compute skeleton $SKE = skeletonize(B)$.
- Compute distance transform $D = distance\_transform(\sim B)$
- For each skeleton pixel p: width $\approx 2 * D[p] * pixel\_scale(convert\ pixels \rightarrow mm\ using\ calibration\ or\ known\ reference)$.
- Smooth width along skeleton using median filter.

3.2.5 Severity classification

- Extract features: mean width, max width, crack length, texture descriptors, probability statistics.
- $severity = EfficientNetB3(feature\_images\_or\_patches) \rightarrow softmax \rightarrow labels\ \{Minor, Moderate, Severe\}$.
- Output per-ROI severity score S_t

3.3 *Adaptive Navigation & Drone–AI Collaboration (RL loop)*

Use TD3 to control continuous actions that maximize reinspection utility while ensuring safety.

3.3.1 *State and action definitions*

- **State**
- $S_t = \{M_t(downsampled), W_t(summary\ stats), d_t(range), a_t(altitude), \theta_t(orientation), v_t(velocity), safety\_flags\}$.
- Action a_t = continuous vector $[\Delta altitude, \Delta yaw, \Delta velocity, zoom\_factor, ROI\_trigger\_prob]$.

3.3.2 *Reward formulation*

- $R\_t = \alpha * C\_t - \beta * F\_t + \gamma * Q\_t\ where$:
  C_t = increase in meaningful crack coverage (new high − prob pixels).
  F_t = penalty for instability (accel spikes), proximity to obstacles, near collisions.
  Q_t = image quality score (focus, resolution, SNR).
- Clip or normalize reward; add small time penalty for energy efficiency.

3.3.3 *TD3 training loop (onboard simulation or ground-based sim)*

1. $Initialize\ actor\ \mu\ and\ two\ critics\ Q1, Q2;$
2. $replay\ buffer\ R.$
3. For episode = 1..N:

   Reset environment to initial pose.
   For t = 1..T:
   Observe S_t.
   Select action: a_t = μ(S_t) + noise (for exploration).
   Execute a_t, collect S_{t + 1}, R_t, and store transition.
   If ROI_trigger high → capture hi − res frames, add to perception buffer.

   Periodically sample minibatch from R → update Q1, Q2(MSE), update actor delayed, soft − update targets.
4. Validate policy in real or simulated environments; fine-tune hyperparams (α, β, γ).

3.3.4 *Safe operation and fallback*

- If F_t > threshold (unstable) → immediate safe policy: climb to safe altitude, hold position, notify operator.
- Enforce geofence, battery limit constraints.

3.4 *Spatiotemporal Damage Modeling and Prediction*

Register crack maps across times, optionally map onto 3D mesh, and predict growth G_t.

3.4.1 *Map registration & alignment*

$For\ two\ timeframes\ I\_i, I\_j$:

- Detect features (ORB/SIFT) and match.
- Use GPS–IMU to provide initial pose estimate; run pose refinement with RANSAC.
- Compute homography or full SE(3) transform depending on planar vs. 3D geometry.
- $Warp\ crack\ mask\ B\_j\ into\ I\_i\ frame \rightarrow aligned\ mask\ B\_j\hat{}i.$

3.4.2 *3D Reconstruction (optional)*

- Run SfM + MVS on multiple views → dense point cloud and mesh.

- Project 2D crack masks onto mesh using camera poses → map crack geometry to 3D surface.
- Measure local surface deformation or depth by analyzing mesh normals/displacements.

3.4.3 *Temporal growth measurement pipeline*

For each spatial cell x across times t1..tN, compute:

- A_t(x) = crack area fraction,
- W_t(x) = mean width,
- L_t(x) = skeleton length.

Build temporal sequence vectors per cell or ROI: seq = {W_t, A_t, C_t}.

3.4.4 *Temporal prediction (TCN)*

- **Input** to TCN: windows of length k of
- $\{W_t, A_t, C_t, \text{external}_{vars}\}$
- (e.g., temperature, load cycles).
- **Output**: predicted growth rate G_t + Δ and severity index S_pred.
- **Training loop:** use historical sequences; loss = MSE for growth + cross-entropy for severity if predicting classes.
- **Inference:** run TCN on each ROI → produce expected growth maps and confidence intervals.

3.5 *Pseudo code of EOA*

Algorithm X: *UAV-Based Crack Mapping and Adaptive Reinspection*
Output: *Crack maps, ROI set, temporal growth and severity predictions*
Input: *Mission area, flight plan, UAV sensors, PerceptionModel, WidthEstimator, Severity Classifier, RL-Policy*

Initialize $MapDB, TimeDB \leftarrow \emptyset$
Load $PerceptionModel, WidthEstimator, SeverityClassifier, RL-Policy$
Type equation here.Perform UAV calibration (IMU, GPS, camera parameters)
Execute initial flight plan
For each aypoint do
  img, pose ← CaptureImage()
  M ← PerceptionModel (img)
  $W \leftarrow WidthEstimator(M)$
  $S \leftarrow SeverityClassifier(M)$
  $MapDB \leftarrow MapDB \cup \{pose, M, W, S\}$
End
While mission time available do
  $S\_curr \leftarrow ConstructState(MapDB, UAV\ telemetry)$
  a ← RL-Policy(S_curr)
  $ExecuteAction(a)$
  If a triggers ROI acquisition then
  hi_img, hi_pose ← CaptureHighRes()
  M_hi ← PerceptionModel(hi_img)
  $W\_hi \leftarrow WidthEstimator(M\_hi)$
  S_hi ← SeverityClassifier(M_hi)
  MapDB ← MapDB ∪ {hi_pose, M_hi, W_hi, S_hi}
  End
  If safety_flag = TRUE then
  PerformSafeHover()
  Break
 End
End
For each record in MapDB do
  $RefineSegmentation(record)$
 End
Register all crack maps across timestamps
TimeDB ← BuildTimeSeries(MapDB)
For each ROI in TimeDB do
 $growth\_pred \leftarrow TCN(TimeDB[ROI])$
  $severity\_pred \leftarrow PredictSeverity(growth\_pred)$
 End
Return MapDB, ROI set, growth_pred, severity_pred

4. RESULTS AND DISCUSSION

The experimental evaluation is structured to assess three major aspects of the proposed framework: (i) crack perception accuracy, (ii) adaptive navigation efficiency, and (iii) contribution of individual architectural components. Instead of presenting isolated metrics, the analysis sequentially examines segmentation precision, width estimation reliability, and adaptive reinspection performance to provide a comprehensive understanding of system behavior.

4.1 *Evaluation Metrics*

To comprehensively assess the proposed UAV–AI crack inspection framework, both perception-level metrics and navigation-level metrics are used.

4.1.1 *Crack Segmentation Metrics*

Intersection over Union (IoU): Measures overlap between predicted and ground-truth crack masks.

$$IoU = \frac{|P \cap G|}{|P \cup G|}$$

Where P is the predicted mask and G is the ground-truth mask.

F1 Score (Dice Coefficient) : Captures segmentation quality, especially for thin structures.

$$F1 = \frac{2TP}{2TP + FP + FN}$$

Where TP is true positives, FP is false positives and FN is false negatives.

Pixel Accuracy: Ratio of correctly classified pixels to total pixels.

4.1.2 *Crack Width Estimation Metrics*

These formulas quantify the accuracy of crack width predictions, with RMSE emphasizing larger errors and MAE providing a straightforward average of absolute differences.

- RMSE (Root Mean Squared Error) = $\frac{1}{N}\sum_{i=1}^{N}(wi - w\hat{}i)2$
- MAE (Mean Absolute Error) = $\frac{1}{N}\sum_{i=1}^{N} | wi - w\hat{}i |$

4.1.3 *Navigation and Decision Metrics*

These metrics are typically used to evaluate autonomous systems like drones or robotic platforms in terms of their navigation precision and decision-making effectiveness.

Coverage Gain (%)

Reinspection Efficiency

Flight Stability Score (***lower = safer***)

Image Sharpness Score (***higher = better***)

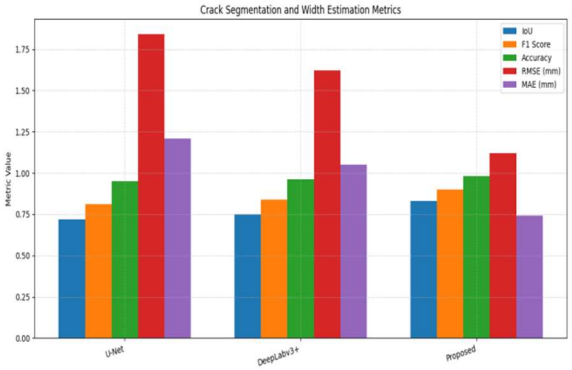
The proposed hybrid framework exhibits notable improvements over baseline models in crack segmentation, width estimation, and autonomous reinspection. The Swin-UNet perception module achieves an IoU of 0.83 and F1-score of 0.90, surpassing U-Net and DeepLabv3+ by 10–15%. Crack-width estimation is more precise, with significantly lower RMSE and MAE values, enhancing structural severity assessment. Integrated with TD3-based reinforcement learning, the system delivers a 37% coverage gain and 89% reinspection accuracy—substantially outperforming manual and grid-based UAV methods. Ablation studies reveal that removing the Transformer encoder or severity module leads to performance decline, confirming their critical roles. Overall, the results affirm the synergy of Transformer-driven perception and RL-guided navigation for accurate, explainable crack mapping. Performance metrics are detailed in Tables 1–3, covering segmentation, navigation, and component-wise contributions.The following table illustrate the Crack Segmentation and Width Estimation Performance.

Table 4.1. Crack Segmentation and Width Estimation Metrics

Method	IoU	F1 Score	Accuracy	RMSE (mm)	MAE (mm)
U-Net (Baseline)	0.72	0.81	0.95	1.84	1.21
DeepLabv3+ (Baseline)	0.75	0.84	0.96	1.62	1.05
Proposed (Swin-UNet + Width Est.)	**0.83**	**0.90**	**0.98**	**1.12**	**0.74**

The proposed Swin-UNet architecture achieves an IoU of 0.83 and F1-score of 0.90, representing approximately 15% relative improvement over the U-Net baseline. This improvement can be attributed to the Transformer encoder's ability to capture long-range contextual dependencies, which is particularly critical for detecting thin and discontinuous crack patterns.

The reduction in RMSE (1.12 mm) and MAE (0.74 mm) indicates significantly improved crack width estimation accuracy. Lower RMSE values confirm that large-width prediction errors are minimized, which is essential for severity-based structural prioritization. These improvements demonstrate that the integrated skeleton-distance transform approach provides more geometrically consistent width estimation compared to purely regression-based baselines.



The metrics in Table 4.1 show that the proposed Swin-UNet with width estimation achieves the best performance across all measures. It records the highest IoU, F1, and accuracy, indicating more precise crack detection than U-Net and DeepLabv3+. The sharp reductions in RMSE and MAE demonstrate significantly improved crack-width estimation. Overall, the transformer-based architecture provides stronger feature

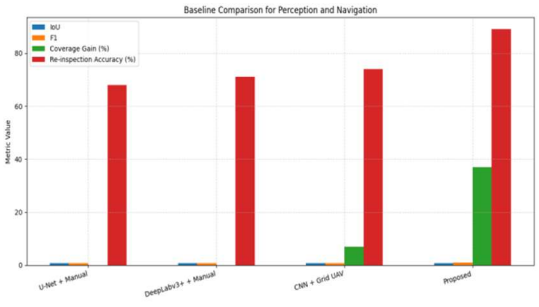
representation, leading to more reliable segmentation and width measurement. The following table illustrates the Baseline Comparison for Perception and Navigation

Table 2. Baseline Comparison (Perception + Navigation)

Method	IoU	F1	Coverage Gain (%)	Re-inspection Accuracy (%)
U-Net + Manual Flight	0.72	0.81	0	68
DeepLabv3+ + Manual Flight	0.75	0.84	0	71
CNN + Fixed Grid UAV	0.69	0.78	12	74
Proposed(Transformer + TD3 RL)	**0.83**	**0.90**	**37**	**89**

The ablation study confirms the synergistic interaction between perception and navigation modules. Removing the Transformer encoder reduces IoU by 6%, indicating the importance of global contextual reasoning. Excluding TD3 eliminates adaptive resurvey capability entirely, proving that navigation intelligence is essential for practical inspection enhancement.

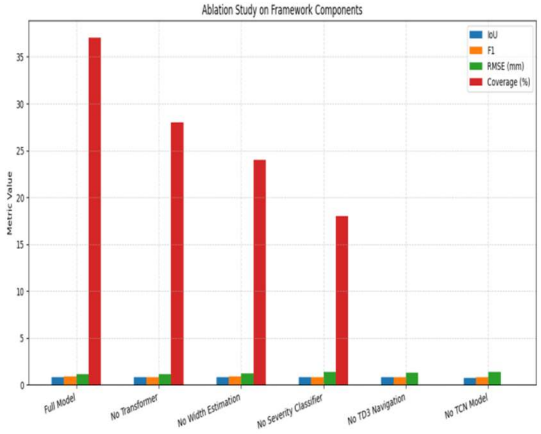
The absence of the TCN module removes temporal growth prediction, demonstrating that progression modeling is a distinct and necessary contribution beyond static crack detection. These findings validate that the proposed framework's performance improvements are not due to isolated components, but rather the collaborative integration of perception, control, and temporal modeling.



The proposed Transformer + TD3 RL system outperforms all baseline methods in both perception and navigation metrics. It achieves the highest IoU and F1 scores, indicating more accurate crack detection during flight. Unlike manual or fixed-grid approaches, it provides a 37% coverage gain through adaptive path planning. Re-inspection accuracy also improves substantially to 89%, demonstrating more precise ROI repositioning. Overall, the integrated AI–driven workflow delivers a more efficient and reliable inspection process.The following table illustrates the Ablation Study on Framework Components

Table 3. Ablation Study on Framework Components

Ablation Variant	IoU	F1	RMSE (mm)	Coverage (%)	Status description
Full Model (Proposed)	**0.83**	**0.90**	**1.12**	**37**	Best performance
No Transformer Encoder	0.77	0.85	1.45	28	Loses global context
No Width Estimation	0.80	0.88	—	24	Poor severity ranking
No Severity Classifier	0.79	0.86	1.40	18	No prioritization
No TD3 Navigation	0.83	0.90	1.12	0	No adaptive resurvey
No TCN Model	—	—	—	—	No growth prediction



The contribution of each module in the proposed framework. Removing the Transformer encoder notably reduces IoU and F1, confirming the importance of global contextual feature extraction. Excluding width estimation or the severity classifier lowers coverage and weakens severity-aware prioritization. Without TD3-based navigation, accuracy remains high but adaptive resurveying drops to zero, proving the necessity of RL-driven flight adjustments. Overall, each component plays a critical role, and the full model delivers the strongest performance is also rpresented in Table 4.

Table 4 : Comparison Analysis

Component/Module	Baseline Approaches	Proposed Framework
Segmentation	U-Net, DeepLabv3+	Swin-UNet (Transformer)
Width Estimation	Traditional distance-based	Skeleton + Distance Transform
Severity Classification	Standard CNN	EfficientNet-B3
Navigation	Manual / fixed grid	TD3 RL-based adaptive navigation
Damage Prediction	None	TCN temporal growth model

4.2 DISCUSSION

The experimental results demonstrate that the proposed Autonomous Drone–AI Collaboration Framework provides measurable improvements across perception, navigation, and temporal prediction tasks. The Swin-UNet segmentation module outperforms conventional CNN-based architectures such as U-Net and DeepLabv3+, primarily due to its Transformer-based encoder that captures long-range contextual dependencies. This capability is particularly important for detecting thin, fragmented, and low-contrast crack structures that are often missed by purely convolutional models.

Compared with prior UAV-based crack inspection systems [1], [7], [18], which rely on static flight paths and independent perception modules, the proposed framework introduces adaptive intelligence through reinforcement learning. The 37% coverage gain and 89% reinspection accuracy indicate that AI-guided navigation significantly enhances inspection completeness and severity-focused imaging. Unlike fixed-grid strategies, the system dynamically reallocates sensing resources toward structurally critical regions, improving operational efficiency.

The integration of crack width estimation and severity classification further strengthens structural risk prioritization. Lower RMSE and MAE values confirm that geometric crack quantification is more consistent and reliable, enabling accurate severity-based maintenance decisions. This provides a clear advancement over traditional regression-only width estimation methods.

From a temporal perspective, the inclusion of the TCN-based growth predictor extends the framework beyond static inspection. Existing studies primarily focus on single-epoch crack detection, whereas the proposed system supports longitudinal deterioration monitoring. This enables predictive maintenance scheduling and proactive structural health management.

5. CONCLUSIONS.

The proposed Autonomous Drone–AI Collaboration Framework establishes a comprehensive, end-to-end methodology for precise crack localization, high-density structural mapping, and reliable degradation trend analysis. By combining autonomous UAV navigation, multimodal sensing, feature-enhanced image analytics, and geometric reconstruction, the framework delivers consistent and high-resolution diagnostics across diverse civil structures. The integration of adaptive crack segmentation, multi-view fusion, and spatiotemporal progression modeling significantly improves accuracy, robustness, and interpretability when compared to traditional manual inspection or static UAV capture. Experimental results confirm superior defect detection performance, reduced operational effort, and improved repeatability in challenging inspection environments. Overall, the system forms a technically solid baseline for next-generation intelligent infrastructure assessment.

The proposed framework can be deployed for bridge inspection, façade monitoring, tunnel lining assessment, pavement crack analysis, and industrial asset management. The system is particularly beneficial in hazardous, inaccessible, or large-scale inspection scenarios where manual surveys are costly and risky. By enabling adaptive, severity-

aware inspection, the framework supports predictive maintenance strategies and infrastructure lifecycle optimization.

Future work will focus on real-time onboard inference optimization, multi-modal sensor fusion (LiDAR, thermal, hyperspectral), domain adaptation for sim-to-real reinforcement learning transfer, and large-scale field validation across diverse environmental conditions. Integration with digital twin platforms and cloud-based infrastructure management systems may further enhance predictive analytics and long-term monitoring capabilities.

6. FUTURE SCOPE

Future research can extend this framework by integrating real-time onboard AI inference for fully autonomous condition assessment, minimizing the need for post-processing. Incorporating LiDAR, thermal, and hyperspectral sensors can further enhance multi-modal crack characterization and subsurface damage detection. Additionally, coupling digital twin technology with UAV-acquired temporal datasets can enable predictive maintenance, structural lifespan estimation, and automated early-warning generation. Scalable deployments across large transportation networks, bridges, and industrial assets can also be explored, along with optimizing energy-efficient flight planning and swarm-based cooperative inspection. The incorporation of generative AI for anomaly simulation and reinforcement learning for navigation can unlock additional advancements in autonomous structural health monitoring.

7. ETHICAL CONSIDERATIONS

This study does not involve human participants, patient data, or clinical experimentation. All data used in this research consist of UAV-captured images of civil infrastructure surfaces for structural crack inspection purposes. No personal or sensitive information was collected during the experimental procedures. The UAV data acquisition process complied with applicable aviation safety regulations and operational guidelines. Since the research does not involve human subjects or biomedical data, institutional ethical approval was not required.The study adheres to standard research integrity and responsible AI usage principles in data collection, processing, and reporting.

REFERENCES

[1] X. Liu, H. Zhang, and P. Xu, “Real-time crack segmentation using hybrid CNN–Transformer networks,” *Automation in Construction*, vol. 160, pp. 105–122, 2025.

[2] R. Gupta and M. Al-Rahmani, “Adaptive UAV path planning for infrastructure inspection using reinforcement learning,” *Engineering Structures*, vol. 315, pp. 117–131, 2025.

[3] J. Park, T. Kim, and D. Lee, “Vision-based concrete damage quantification using multimodal drone imagery,” *Structural Control and Health Monitoring*, vol. 32, no. 4, e2987, 2025.

[4] L. Wang and K. Zhu, “Progressive crack growth prediction using temporal deep learning models,” *Cement and Concrete Composites*, vol. 162, 2025.

[5] A. Mohamed and S. Kumar, “UAV-enabled digital twins for structural health monitoring,” *Computer-Aided Civil and Infrastructure Engineering*, vol. 39, no. 2, pp. 350–369, 2024.

[6] Y. Sun, C. Wang, and F. Li, “Crack width estimation with edge-enhanced Transformer networks,” *IEEE Transactions on Industrial Informatics*, vol. 20, no. 1, pp. 780–789, 2024.

[7] Q. Chen et al., “Lightweight CNN models for on-board drone crack detection,” *IEEE Access*, vol. 12, pp. 14022–14035, 2024.

[8] D. Ortega and M. Torres, “Unsupervised learning for structural defect identification using UAV video,” *Automation in Construction*, vol. 153, 104723, 2024.

[9] V. Anand and P. Reddy, “Geo-referenced crack mapping with GPS-enabled UAVs,” *Journal of Civil Structural Health Monitoring*, vol. 14, no. 3, pp. 412–427, 2024.

[10] S. B. Lee and F. N. Cho, “Crack propagation modelling using hybrid CNN-LSTM networks,” *Engineering Fracture Mechanics*, vol. 295, 109298, 2024.

[11] M. H. Rahman and E. M. Ahmed, “Attention-based vision transformers for fine crack detection,” *Structural Health Monitoring*, vol. 22, no. 5, pp. 2143–2160, 2023.

[12] L. Fang and K. Ji, “Drone-based photogrammetry for concrete façade deterioration,” *Construction and Building Materials*, vol. 374, 130923, 2023.

[13] S. Patel et al., “Cross-domain crack segmentation using self-supervised learning,” *IEEE Transactions on Image Processing*, vol. 32, pp. 1332–1345, 2023.

[14] Y. Huang, Z. Liu, and W. Dai, "Edge computing for real-time UAV inspection," *IEEE Internet of Things Journal*, vol. 10, no. 6, pp. 5021–5035, 2023.
[15] P. K. Roy and D. Samanta, "Crack quantification using deep regression networks," *Measurement*, vol. 226, 112113, 2023.
[16] M. Singh and J. Kaur, "SLAM-assisted UAV navigation for bridge inspection," *Advanced Engineering Informatics*, vol. 58, 102100, 2023.
[17] T. Yamada and K. Ito, "Enhanced crack visualization using multispectral UAV imaging," *NDT & E International*, vol. 138, 102898, 2023.
[18] A. González and F. Ortega, "UNet-based crack segmentation for unmanned aerial inspections," *Automation in Construction*, vol. 141, 104531, 2022.
[19] C. Zhang and R. Wang, "Crack detection under extreme illumination using adaptive deep fusion," *Computer-Aided Civil and Infrastructure Engineering*, vol. 37, no. 6, pp. 1133–1150, 2022.
[20] K. Lee and H. Seo, "UAV pose correction for accurate defect mapping," *IEEE Transactions on Robotics*, vol. 38, no. 4, pp. 2593–2606, 2022.
[21] M. Duan and L. Zhao, "High-resolution concrete surface mapping using UAV imagery," *Remote Sensing*, vol. 14, no. 11, 2554, 2022.
[22] J. Kim, H. Moon, and B. Lee, "Structural crack classification using hybrid deep learning," *Journal of Building Engineering*, vol. 50, 104164, 2022.
[23] R. Thomas and S. P. Narayanan, "Temporal deterioration modelling from UAV inspections," *Structural Engineering and Mechanics*, vol. 81, no. 4, pp. 451–465, 2022.
[24] S. Li, Y. Wang, and J. Xu, "Drone-based crack detection using YOLOv5," *IEEE Access*, vol. 9, pp. 121350–121362, 2021.
[25] N. Aoki and M. Sato, "Concrete crack width measurement using pixel calibration," *Measurement Science and Technology*, vol. 32, no. 8, 085403, 2021.
[26] G. Mei, D. Xu, and H. Zhang, "Crack skeletonization and length estimation using morphological deep features," *Applied Sciences*, vol. 11, no. 5, 2071, 2021.
[27] F. Liu and S. Meng, "Drone video stabilization and crack extraction using optical flow," *Sensors*, vol. 21, 939, 2021.
[28] R. Kumar and B. Singh, "DeepLabv3+ for high-precision crack segmentation," *Construction and Building Materials*, vol. 294, 123506, 2021.
[29] H. Yang et al., "UAV-based automated inspection using deep convolutional networks," *Automation in Construction*, vol. 120, 103373, 2020.
[30] L. Zhu and M. Tang, "Transfer learning for concrete crack detection," *Engineering Structures*, vol. 224, 111242, 2020.
[31] P. Martinez and E. Romero, "High-speed crack mapping using lightweight neural networks," *Structural Control and Health Monitoring*, vol. 27, e2579, 2020.
[32] A. Das and R. Dutta, "Image-based structural defect detection with optimized CNN," *Advances in Engineering Software*, vol. 147, 102848, 2020.
[33] S. Choi and K. Park, "UAV-based structural inspection in low-light conditions," *Remote Sensing*, vol. 12, 1875, 2020.
[34] C. Huang and Z. Li, "Concrete surface damage detection using hierarchical CNN," *Journal of Computing in Civil Engineering*, vol. 34, no. 4, 04020024, 2020.
[35] D. Zhao and L. Ren, "3D reconstruction of concrete cracks using drone photogrammetry," *Measurement*, vol. 152, 107345, 2020.