

AURA: An AI Powered Universal Augmented Reality Digital Twin Platform

Mrs. Haseenath Jaseela A, Kishore M, Beno shine A, Vishak V, Hurdhay N Nair, Abin Meshack.R

Assistant professor, students Department of AI & DS, Loyola Institute of Technology and
Science, Thovalai, Kanyakumari Dist, Tamil Nadu.

Email:haseena.jas84@gmail.com.

ABSTRACT

Contemporary digital twin deployments remain domain locked, capture expensive, and largely non interactive: they are engineered separately for manufacturing, smart city, healthcare, or heritage use cases, depend on specialist LiDAR or CAD pipelines, refresh on batch schedules rather than continuously, and expose their analytics as opaque dashboards with no augmented reality surface, no explanation, and no multilingual or accessibility support. This paper presents AURA, an AI powered universal augmented reality digital twin platform that unifies commodity device spatial capture, neural 3D reconstruction, live state synchronisation, predictive simulation, explainable AI (XAI), and a multilingual AR assistant (Tamil, English, Hindi) within a single, domain agnostic modular architecture. The system is organised into four cooperating modules Spatial Capture and Twin Construction, Live Synchronisation and Perception, Predictive Simulation and Explainability, and Immersive Interaction and Accessibility coordinated through a Central Twin Engine and a Universal Twin Knowledge Graph whose schema is entity typed rather than domain typed, allowing the same pipeline to instantiate a factory floor, a campus, a heritage monument, or a hospital ward without re engineering. We describe the system architecture, the algorithmic foundation of each module (visual inertial SLAM for AR registration, neural radiance and Gaussian splat reconstruction, CNN semantic segmentation, Kalman filter multi sensor fusion, LSTM state forecasting, Random Forest anomaly classification, and post hoc feature attribution explanation), and the end to end workflow from scan ingestion to feedback driven retraining. We further present a comparative analysis against existing digital twin and AR platforms across fifteen parameters and outline an experimental protocol for evaluating reconstruction fidelity, registration accuracy, forecasting error, anomaly classification performance, and end to end latency, and report expected performance ranges pending full scale empirical evaluation. The proposed framework demonstrates how universality, explainability, real time synchronisation, and inclusive interaction can be combined to deliver transparent, device agnostic, context aware spatial intelligence, and it establishes a foundation for future extensions involving generative scene authoring, multi user shared twins, and smart city scale federation.

1. INTRODUCTION

Physical environments are increasingly instrumented. Factories, campuses, transport corridors, hospitals, and heritage sites now generate continuous streams of sensor, camera, and usage data, and organisations are turning to digital twins synchronised virtual replicas of physical assets to interpret that data, simulate interventions, and plan maintenance. In parallel, consumer augmented reality has matured: modern smartphones and headsets ship with inertial measurement units, depth sensors, and on device tracking frameworks capable of stable six degree of freedom pose estimation.

Despite this convergence, the two capabilities are rarely delivered together. Digital twin platforms remain enterprise dashboards consumed on flat screens, while AR applications remain visually rich but analytically shallow overlays disconnected from any authoritative model of state. Where the two are combined, the resulting systems are bespoke: a twin built for predictive maintenance on an assembly line cannot be redeployed to a museum or a ward without rewriting its schema, its ingestion layer, and its analytics. The cost of this fragmentation falls hardest on exactly the organisations least able to absorb it regional institutions, municipal bodies, and small enterprises.

This paper proposes AURA, an AI powered universal augmented reality digital twin platform that integrates neural spatial reconstruction, real time sensor synchronisation, predictive simulation, explainable analytics, and multilingual immersive interaction within a single modular architecture. Rather than treating these capabilities as isolated features, AURA is designed as a research contribution centred on architectural universality, algorithmic justification, and measurable performance gains over existing systems. The remainder of this paper reviews existing systems and their limitations, presents the proposed system and identifies the research gap, describes the algorithmic foundation, details the system workflow and layered architecture, explains the novel contributions, outlines the experimental setup and expected performance, presents a comparative analysis, discusses advantages and future scope, and concludes the paper.

2. EXISTING SYSTEM

Existing digital twin and augmented reality platforms, including commercial industrial suites and academic prototypes, exhibit recurring limitations across five related technology categories.

- **Conventional 3D Visualization:** existing 3D systems support rotation, zoom, pan, camera and object inspection, but generally lack contextual AI and predictive assistance.
- **Augmented Reality:** AR provides immersive visualization of digital content in physical space, but often has limited natural language and intelligent interaction.
- **AI Assistants:** AI systems support question answering and conversational interaction, but may not understand the selected 3D component, its spatial position, or its Digital Twin context.
- **Computer Vision:** CV enables object recognition, semantic understanding and gesture interaction, but does not itself provide a complete Digital Twin with live state, simulation and prediction.
- **Digital Twin Platforms:** these systems focus on visualization, monitoring, simulation and data management, while conversational, multimodal and accessibility aware interfaces remain limited.

Major Limitations

The reviewed approaches commonly show: (i) domain locked architectures; (ii) expensive LiDAR/CAD/BIM based capture requirements; (iii) static or batch synchronization instead of continuous updates; (iv) visualization without future state prediction; (v) opaque analytics with limited explanation; (vi) AR registration and pose drift challenges; (vii) limited universality across commodity devices; (viii) monolingual interfaces and weak accessibility support; and (ix) closed schemas that restrict interoperability and third party extension.

3. PROPOSED SYSTEM

AURA is organized into four cooperating modules, coordinated by a Central Twin Engine and a Universal Twin Knowledge Graph, as illustrated in Fig. 1. The Central Twin Engine mediates all cross module communication: it receives spatial and semantic information from the Spatial Capture and Twin Construction module, perception and state information from the Live Synchronisation and Perception module, predicted states and explanations from the Predictive Simulation and Explainability module, and interaction and accessibility context from the Immersive Interaction and Accessibility module, fusing them into a unified and explainable digital twin response. Each module is decomposed below into its constituent components and its interaction points with the remaining modules.

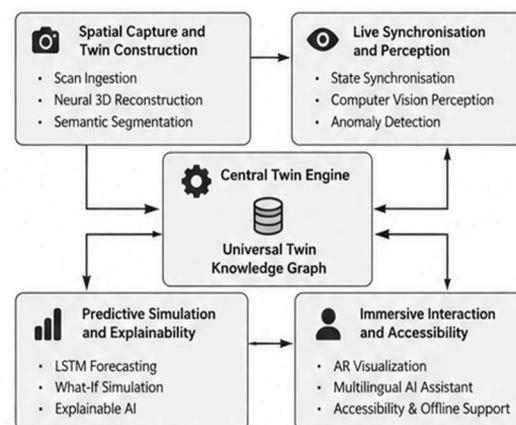


Fig. 1. Framework diagram of the AURA proposed system: four cooperating modules coordinated by a Central Twin Engine and Universal Twin Knowledge Graph.

A. Spatial Capture and Twin Construction Module

This module converts physical space captures obtained from commodity devices into a structured and interactive digital twin and is composed of three tightly coupled components. (1) The Commodity Device Scan Ingestion component processes captured image sequences and estimates camera trajectories using visual inertial odometry for accurate spatial registration. (2) The Neural 3D Reconstruction component generates a detailed three dimensional representation of the captured environment using Neural Radiance Fields (NeRF) and 3D Gaussian Splatting, enabling photorealistic and interactive visualization without requiring dedicated scanning equipment or CAD models. (3) The Semantic Segmentation component applies CNN based scene understanding to identify meaningful objects and regions and converts them into structured entities within the Universal Twin Knowledge Graph. Interaction: the reconstructed geometry and semantic entities are forwarded to the Live Synchronisation and Perception module for state interpretation and to the Immersive Interaction and Accessibility module for AR based visualization and interaction.

B. Live Synchronisation and Perception Module

This module maintains consistency between the physical environment and its digital representation while extracting meaningful visual and contextual information. It is composed of three components. (1) The State Synchronisation component updates the digital twin representation based on available system state information and maintains the current condition of represented entities. (2) The Computer Vision Perception component performs object recognition and semantic understanding to identify changes and relevant entities within the environment. (3) The Anomaly Detection component applies Random Forest classification to identify abnormal or unexpected states from the available feature information. Interaction: perception and anomaly results are supplied to the Predictive Simulation and Explainability module for future state analysis, while updated entity states and semantic information are

provided to the Immersive Interaction and Accessibility module for contextual AR rendering.

C. Predictive Simulation and Explainability Module

This module provides predictive intelligence and interpretable analysis over the digital twin and is composed of three components. (1) The LSTM Based State Forecasting component analyses historical state sequences to predict future conditions of represented entities. (2) The What If Scenario Simulation component evaluates alternative conditions specified by the user and estimates their possible effects on the digital twin state. (3) The Explainable AI component generates human readable explanations for predictions, detected anomalies, and recommended actions by identifying the major contributing factors behind each output. Interaction: this module receives synchronised states and anomaly information from the Live Synchronisation and Perception module and forwards predicted results, explanations, and scenario outcomes to the Immersive Interaction and Accessibility module for presentation to the user.

D. Immersive Interaction and Accessibility Module

This module provides the user facing interface through which users interact with the digital twin and access its intelligent capabilities. It is composed of three components. (1) The AR Visualization and Registration component overlays digital twin information onto the physical environment using visual inertial registration, enabling users to inspect objects and spatial entities directly. (2) The Multilingual AI Assistant supports Tamil, English, and Hindi for text and voice based interaction, allowing users to query the digital twin and receive contextual information. (3) The Accessibility and Offline Support component provides audio descriptions, reduced motion visualization, high contrast rendering, haptic alternatives, and locally cached twin information for constrained environments. Interaction: this module receives reconstructed geometry, semantic information, predicted states, anomaly results, and explanations from the other modules and

presents them through AR overlays, text, voice, and accessible interaction modes.

E. Inter Module Interaction

The four modules are not independent pipelines but form a coordinated system in which each module produces information consumed by other modules through the Central Twin Engine. Fig. 2 summarizes the resulting data flow: captured spatial information flows into reconstruction and semantic understanding; reconstructed entities flow into perception and synchronisation; identified states and anomalies flow into prediction and scenario simulation; predicted results and explanations flow into immersive interaction; and user feedback and accessibility context are returned to the Central Twin Engine to refine subsequent processing. The Universal Twin Knowledge Graph maintains the shared representation of entities, relationships, spatial information, metadata, and system state, ensuring that all modules operate using consistent contextual information. This coordinated architecture enables AURA to provide a unified, explainable, device agnostic, and accessible augmented reality digital twin experience.

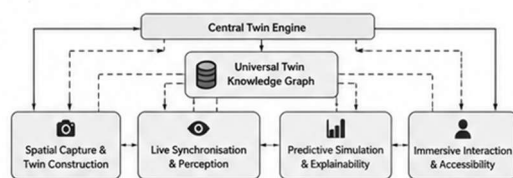


Fig. 2. Module interaction diagram: spatial capture, perception, prediction, explainability, and immersive interaction coordinated through the Central Twin Engine and Universal Twin Knowledge Graph.

4. RESEARCH GAP

Research has progressed along separate tracks including neural reconstruction and SLAM, Digital Twin analytics, and AR/spatial interaction. Neural reconstruction can generate detailed geometry, Digital Twin analytics can model state and prediction, and AR can improve spatial presentation; however, their unified integration remains limited. A 3D model may provide visualization without AI context, an AI assistant may lack awareness of the physical scene, AR may lack contextual intelligence, and Digital Twin platforms may

lack natural conversational interaction. Therefore, there is a need for a unified, explainable, device agnostic and accessibility aware AR Digital Twin framework that connects the physical environment, Digital Twin, AI, perception, prediction and user interaction the gap AURA is designed to address.

5. ALGORITHM

Visual Inertial Odometry and Pose Estimation

Purpose: to recover six degree of freedom camera pose for scan ingestion and AR overlay registration. **Working Principle:** visual feature tracks and inertial pre integration are jointly optimised by minimising a combined reprojection and inertial residual over a sliding window of keyframes, with state propagation and correction of the form $\hat{x}(k|k) = \hat{x}(k|k-1) + K(k)(z(k) - H(k)\hat{x}(k|k-1))$, where $K(k)$ is the Kalman gain. **Advantages:** robust to short visual occlusion, drift bounded through loop closure, and runnable on commodity mobile hardware. **Why Selected:** registration accuracy is the binding constraint on AR usefulness, and a tightly coupled visual inertial formulation gives the metric scale that monocular only tracking cannot recover.

Neural 3D Reconstruction

Purpose: to construct photorealistic, novel view renderable twin geometry from commodity captures. **Working Principle:** a continuous scene function maps position and view direction to colour and density, with pixel colour obtained by volume rendering along the camera ray, $C(r) = \sum T(i)(1 - \exp(-\sigma(i)\delta(i)))c(i)$; an explicit Gaussian splat variant replaces the implicit field with rasterised anisotropic primitives, trading memory for interactive frame rates. **Advantages:** high fidelity from unstructured photographs with no scanner or CAD prerequisite. **Why Selected:** neural and splat based representations approach survey capture quality at a fraction of the acquisition cost, directly addressing the capture cost limitation of existing systems.

CNN Semantic Segmentation

Purpose: to convert undifferentiated reconstructed geometry into typed, queryable twin entities. **Working Principle:** an encoder decoder convolutional network assigns each

pixel and, after back projection, each surface region a class label, trained under a cross entropy objective and evaluated by intersection over union, $\text{IoU}(c) = |\text{P}(c) \cap \text{G}(c)| / |\text{P}(c) \cup \text{G}(c)|$. Advantages: transfer learning from pretrained backbones sharply reduces the labelled data required per new deployment. Why Selected: the same segmentation stage populates a domain agnostic graph schema whether the scene is a workshop or a gallery, which is what makes the twin universal.

Multi Sensor Fusion (Kalman Filter)

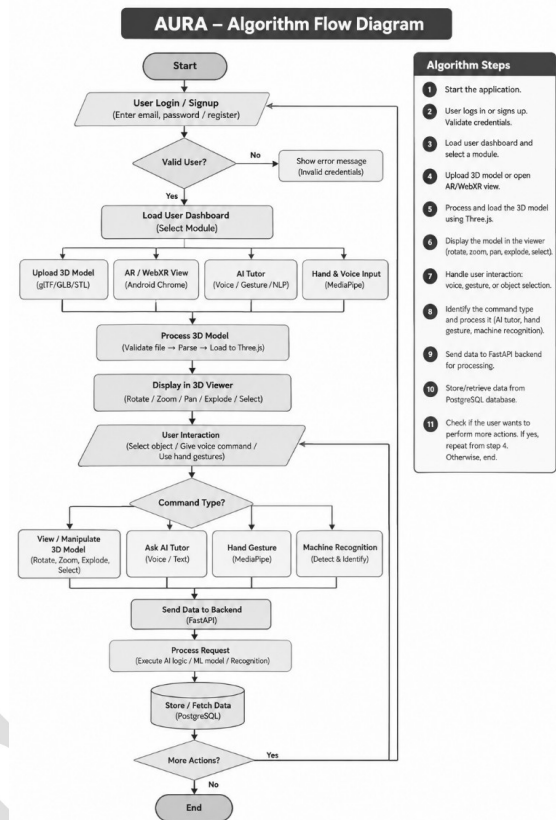
Purpose: to reconcile noisy, asynchronous local sensor streams into one coherent per entity state estimate. Working Principle: the state is propagated and corrected recursively via the standard Kalman predict update recursion. Advantages: optimal under linear Gaussian assumptions, computationally light, and naturally tolerant of missing samples. Why Selected: sensor dropout is the norm rather than the exception in field deployments, and recursive estimation keeps the twin synchronised without blocking on complete observations.

Random Forest Anomaly Classification

Purpose: to classify fused state windows into normal or typed anomaly categories. Working Principle: an ensemble of decision trees is trained via bootstrap aggregation, with the final prediction obtained by majority voting, $\hat{y} = \text{mode}\{h_1(x), h_2(x), \dots, h_T(x)\}$. Advantages: robust to overfitting, handles mixed categorical and numerical features, and yields feature importance scores usable for explanation. Why Selected: the feature importance output aligns naturally with AURA's explainability requirement.

Recurrent State Forecasting (LSTM)

Purpose: to project entity state forward for predictive maintenance and what if simulation. Working Principle: an LSTM cell maintains a gated memory across the fused state sequence via gate equations such as $f(t) = \sigma(W_f[h(t-1), x(t)] + b_f)$, with forecasts evaluated by root mean square error over the horizon, RM



$SE = \sqrt{(1/H)\sum(\hat{y}_h - y_h)^2}$. Advantages: captures long range temporal dependence that window based regressors miss. Why Selected: degradation and occupancy patterns are strongly autocorrelated over hours to weeks, a regime where gated recurrence remains competitive at modest data scale.

Explainable AI (XAI)

Purpose: to generate human readable justifications for anomaly flags and forecasts. Working Principle: post hoc, model agnostic feature attribution assigns each input an additive contribution to the prediction, $g(z') = \phi_0 + \sum \phi_j \cdot z'_j$; the top ranked ϕ_j are surfaced directly in the AR overlay. Advantages: increases operator trust and supports the 'right to explanation' emphasised in trustworthy AI literature. Why Selected: an alert anchored to a physical machine is actionable only if the operator can see why it fired.

Multilingual NLP and Speech Pipeline

Purpose: to power the multilingual assistant, spatial query understanding, and voice interaction. Working Principle: text is tokenised and embedded using multilingual transformer representations, followed by intent classification and slot extraction against the twin graph; acoustic to text models transcribe

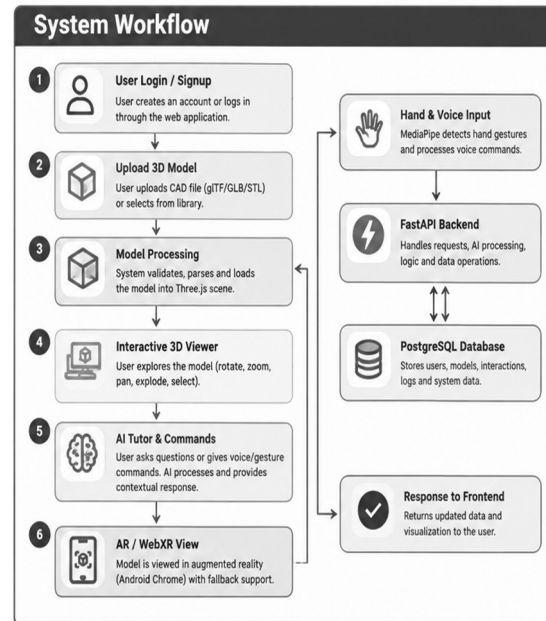
spoken queries and text to speech synthesises replies in the user's chosen language. Advantages: shared multilingual representations allow one model to serve Tamil, Hindi, and English without separate per language pipelines.

Twin Alert Ranking

Purpose: to merge heterogeneous signals into a single ordered set of overlays so the AR view surfaces what matters. Working Principle: a weighted linear combination ranks candidate entities, $\text{Score} = w1 \cdot \text{Anomaly} + w2 \cdot \text{Urgency} + w3 \cdot \text{Proximity} - w4 \cdot \text{Uncertainty}$, where Proximity derives from user pose and Uncertainty from forecast variance and twin fidelity divergence. The linear form keeps the ranking explainable under the same attribution machinery used for XAI. Evaluation metrics such as Precision, Recall and F1 Score are formalised in the standard way from true positives, false positives and false negatives, and registration quality is measured by absolute trajectory error, $\text{ATE} = \sqrt{\frac{1}{N} \sum (t(i, \text{est}) - t(i, \text{gt}))^2}$.

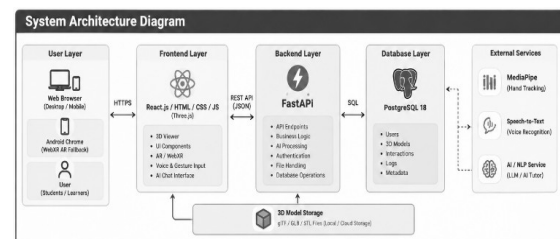
6. SYSTEM WORKFLOW

The end to end workflow traces the AURA system from onboarding to feedback driven retraining. Following registration and selection of a deployment domain, the user performs a commodity device scan; frames are quality filtered and posed, a neural reconstruction is fitted, and segmented entities are written into the knowledge graph. Sensor streams are then bound to those entities, after which the twin enters its steady state loop: fuse, detect, forecast, explain, rank, render. The multilingual assistant and accessibility layer remain available throughout the session, and offline mode is reachable at any stage. Operator corrections – a mislabelled entity, a dismissed false alarm, a confirmed fault – are collected and fed back into model retraining, while the fidelity monitor independently triggers partial re scans as the physical space changes. All metrics are evaluated over the held out test split described in the Experimental Setup.



7. SYSTEM ARCHITECTURE

AURA follows a layered architecture spanning five layers. The Client layer captures user input (camera frames, pose, voice, touch) on phones, tablets, and headsets and forwards it to the Twin Engine, which orchestrates calls to the Reconstruction (NeRF/3DGS, VIO), Semantics (CNN segmentation), Synchronisation and Prediction (Kalman filter, Random Forest, LSTM, XAI), and Interaction (NLP, STT/TTS, accessibility) services. All services read from and write to the shared Universal Twin Knowledge Graph and time series store, and consolidated, explained results are rendered back through the Presentation layer as registered AR overlays, speech, or translated text.



9. NOVEL CONTRIBUTION

AURA's novelty lies not in any single algorithm but in their integration:

- A Universal Twin Schema that is entity typed rather than domain typed, allowing one pipeline to instantiate factory, campus, ward, or heritage twins without re engineering.
- Commodity Device Twin Construction, replacing scanner and CAD dependent capture

with neural reconstruction from ordinary phone video.

- Explainability Applied Uniformly across anomaly, forecast, and overlay ranking outputs rather than to a single model in isolation.
- Continuous Fidelity Monitoring that treats twin reality divergence as a first class measured quantity with an automatic re scan response.
- A Multilingual AR Assistant spanning chat, voice, and live translation for Tamil, English, and Hindi, delivered over spatial rather than tabular context.
- Spatially Aware Alert Ranking that fuses anomaly severity, urgency, user proximity, and uncertainty into one interpretable score.
- Accessibility as a Core Module audio scene description, reduced motion rendering, and haptic substitution rather than an afterthought.

8. EXPERIMENTAL SETUP

Dataset

As AURA is at the design and partial implementation stage, this section specifies a proposed dataset rather than reporting results from a fully deployed system. The proposed dataset comprises: (i) a scene corpus of approximately 40 indoor and outdoor scenes across at least three domains (an academic laboratory block, a heritage site, and a light manufacturing workshop), each with 300 800 posed frames from commodity smartphones, with 8 scenes additionally captured on survey equipment for reconstruction ground truth; (ii) a segmentation set of approximately 12,000 annotated frames spanning 30 domain agnostic entity classes; (iii) a sensor data corpus of at least 6 months of temperature, vibration, occupancy, power, and air quality readings, combining public industrial monitoring datasets with pilot deployments; (iv) an incident/anomaly table of labelled fault and deviation events, augmented with synthetically injected anomalies; and (v) a query corpus of approximately 6,000 spatial queries in Tamil, English, and Hindi with intent and entity slot labels. Data will be partitioned 70/15/15 into training, validation, and test splits, stratified by domain and scene.

Hardware and Software

Development and training are planned on a workstation with an NVIDIA RTX 3060 (12 GB VRAM) or an equivalent cloud GPU

instance, 32 GB RAM, and a multi core CPU (Intel i7 class or equivalent), running Ubuntu 22.04; on device inference will be validated on mid range Android handsets with ARCore class tracking support. The software stack comprises Python 3.10, PyTorch 2.x for reconstruction, segmentation, and sequence components, scikit learn for Random Forest classification, HuggingFace Transformers for the multilingual NLP pipeline, Open3D and COLMAP for geometry handling and pose bootstrapping, Unity or WebXR for the AR client, and FastAPI with a PostgreSQL/TimescaleDB backend for orchestration and storage.

Training and Testing

Training follows a per module protocol appropriate to each architecture. Neural reconstruction will be fitted per scene for 20,000 30,000 iterations with the Adam optimiser (learning rate $5e-4$, exponential decay), with the Gaussian splat variant trained for 30,000 iterations under adaptive density control. The segmentation network will use transfer learning from a pretrained backbone (ResNet 50 or EfficientNet B0), fine tuned for 40 epochs with batch size 16. The Random Forest anomaly classifier will be trained with 200 trees and 5 fold cross validation. The LSTM forecaster will use two stacked layers of 128 hidden units, sequence length 96, trained for 60 epochs with early stopping. The multilingual intent classifier will be fine tuned from a pretrained multilingual transformer (mBERT or XLM R) for 10 epochs. All models will be evaluated on the held out 15% test split.

Evaluation Metrics

The system will be evaluated using reconstruction PSNR, segmentation mIoU, registration ATE and drift over session duration, anomaly classification Accuracy/Precision/Recall/F1, forecast RMSE/MAPE, end to end overlay latency, sustained client frame rate, explanation usefulness (operator rating), and task completion time relative to a dashboard baseline.

10. EXPECTED PERFORMANCE ANALYSIS

Full scale empirical evaluation is in progress; this section identifies the metrics that will be reported rather than measured results, so that

the evaluation protocol can be assessed independently of final numbers. Target ranges will be informed by reported results for structurally similar components in prior neural reconstruction, SLAM, Random Forest, and sequence forecasting studies, and represent goals for the completed system rather than claims of achieved performance. Table 1 lists the metrics that will be populated with measured values upon completion of the evaluation described in the Experimental Setup.

TABLE 1. EXPECTED / TARGET PERFORMANCE METRICS

Evaluation Metric	What It Measures
Reconstruction PSNR	Fidelity of neural 3D reconstruction against held out views
Segmentation mIoU	Accuracy of CNN semantic segmentation
Registration ATE	AR overlay pose accuracy and drift over session duration
Accuracy / Precision / Recall F1	Random Forest anomaly classification performance
Forecast RMSE / MAPE	LSTM state forecasting error over the prediction horizon
End to end overlay latency	Time from sensor update to rendered, explained overlay
Sustained client frame rate	Real time responsiveness of the AR client
Explanation usefulness (operator rating)	Perceived value of XAI outputs
Task completion time	Efficiency relative to a conventional dashboard baseline

11. COMPARATIVE ANALYSIS

TABLE 2. EXISTING DIGITAL TWIN / AR SYSTEMS VS. AURA

Parameter	Existing Systems	AURA
Domain coverage	Single vertical, hard coded	Universal entity typed schema
Capture requirement	LiDAR / CAD / BIM	Commodity phone capture
Reconstruction method	Manual modelling / photogrammetry	Neural (NeRF / 3D Gaussian Splatting)
Synchronisation	Batch or periodic refresh	Continuous Kalman fused state
Twin fidelity tracking	Not measured	Monitored with automatic re scan trigger
Prediction	Visualisation only / rare	LSTM forecasting + what if simulation
Explainability	None (opaque alerts)	Built in XAI layer on all outputs
AR interface	Absent or drift prone	VIO registered, drift corrected overlays
Alert prioritisation	Flat alert list	Proximity and uncertainty aware ranking
Multilingual support	English only (typical)	Tamil, English, Hindi
Voice interaction	Rare	Integrated multilingual voice assistant

Accessibility support	Minimal	Dedicated core module
Offline operation	Not supported	Cached twin slice for field use
Feedback driven retraining	Rare	Supported
Architecture	Siloed / single purpose	Unified modular framework

12. ADVANTAGES

- Domain portable deployment through a universal, entity typed twin schema.
- Dramatically lower capture cost via commodity device neural reconstruction.
- Transparent, trust building alerts and forecasts through integrated explainability.
- Genuine live correspondence between model and reality, with divergence measured rather than assumed.
- Actionable foresight through predictive simulation and counterfactual what if analysis.
- Reduced language barriers via native multilingual chat and voice support.
- More inclusive spatial experiences through dedicated accessibility provision.
- Resilience in low connectivity field conditions via the cached offline twin slice.

13. FUTURE SCOPE

Planned extensions include: Generative Scene Authoring, in which a language model proposes and instantiates twin entities and layout variants directly from natural language description; Multi User Shared Twins supporting co located and remote collaborators in one registered session; Federated Twin Interchange, allowing independently built twins to exchange entities across an open schema without centralising raw data; Edge and On Device Inference to move segmentation and anomaly detection onto the handset for latency and privacy sensitive deployments incorporating personal and ambient sensing; Blockchain Anchored Provenance for verifiable inspection and maintenance records; and Smart City Scale Federation linking AURA instances with municipal transport, utility, and emergency data feeds.

14. CONCLUSION

This paper's principal contribution is architectural rather than algorithmic: AURA demonstrates a design pattern for integrating

commodity device neural reconstruction, continuous state synchronisation, explainable prediction, and inclusive augmented reality interaction into a single coordinated spatial intelligence pipeline an integration that, per the identified research gap, does not appear jointly addressed in the surveyed literature. Concretely, the framework contributes: (1) a universal, entity typed twin schema and module interaction architecture generalisable beyond any one vertical; (2) a ranking formulation that fuses anomaly severity, urgency, proximity, and uncertainty into a single explainable score, providing a template for combining heterogeneous AI outputs under one interpretable objective; (3) an explicit twin fidelity mechanism that treats model reality divergence as a measured, actionable quantity; and (4) an experimental protocol and comparative framework that establish measurable criteria against which this and future systems can be evaluated. Beyond aggregating features, this positions AURA as a research contribution to the design of transparent, inclusive, multi modal spatial recommender and monitoring architectures, providing a concrete, measurable path toward generative authoring, multi user twins, and smart city federation.

REFERENCES

- [1] M. Grieves and J. Vickers, "Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems," in *Transdisciplinary Perspectives on Complex Systems*, Springer, 2017, pp. 85–113.
- [2] F. Tao, H. Zhang, A. Liu, and A. Y. C. Nee, "Digital twin in industry: State of the art," *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 2405–2415, 2019.
- [3] A. Fuller, Z. Fan, C. Day, and C. Barlow, "Digital twin: Enabling technologies, challenges and open research," *IEEE Access*, vol. 8, pp. 108952–108971, 2020.
- [4] C. Zhuang, J. Liu, and H. Xiong, "Digital twin based smart production management and control framework for the complex product assembly shop floor," *Int. J. Adv. Manuf. Technol.*, vol. 96, pp. 1149–1163, 2018.
- [5] R. Mur Artal and J. D. Tardos, "ORB SLAM2: An open source SLAM system for monocular, stereo, and RGB D cameras," *IEEE Trans. Robot.*, vol. 33, no. 5, pp. 1255–1262, 2017.
- [6] T. Qin, P. Li, and S. Shen, "VINS Mono: A robust and versatile monocular visual inertial state estimator," *IEEE Trans. Robot.*, vol. 34, no. 4, pp. 1004–1020, 2018.
- [7] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "NeRF: Representing scenes as neural radiance fields for view synthesis," in *Proc. ECCV*, 2020.
- [8] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, "3D Gaussian splatting for real time radiance field rendering," *ACM Trans. Graph.*, vol. 42, no. 4, 2023.
- [9] O. Ronneberger, P. Fischer, and T. Brox, "U Net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*, 2015, pp. 234–241.
- [10] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, 2018.
- [11] R. E. Kalman, "A new approach to linear filtering and prediction problems," *J. Basic Eng.*, vol. 82, no. 1, pp. 35–45, 1960.
- [12] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [13] T. P. Carvalho, F. A. A. M. N. Soares, R. Vita, R. P. Francisco, J. P. Basto, and S. G. S. Alcázar, "A systematic literature review of machine learning methods applied to predictive maintenance," *Comput. Ind. Eng.*, vol. 137, 2019.
- [14] S. Hochreiter and J. Schmidhuber, "Long short term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [15] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in *Proc. NIPS*, 2017, pp. 4765–4774.
- [16] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you?: Explaining the predictions of any classifier," in *Proc. ACM SIGKDD*, 2016, pp. 1135–1144.
- [17] A. Conneau et al., "Unsupervised cross lingual representation learning at scale," in *Proc. ACL*, 2020, pp. 8440–8451.
- [18] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre training of deep bidirectional transformers for language understanding," in *Proc. NAACL HLT*, 2019.
- [19] J. L. Schonberger and J. M. Frahm, "Structure from motion revisited," in *Proc. IEEE CVPR*, 2016, pp. 4104–4113.
- [20] T. Müller, A. Evans, C. Schied, and A. Keller, "Instant neural graphics primitives with a multiresolution hash encoding," *ACM Trans. Graph.*, vol. 41, no. 4, 2022.
- [21] A. Dai, A. X. Chang, M. Savva, M. Halber, T. Funkhouser, and M. Niessner, "ScanNet: Richly annotated 3D reconstructions of indoor scenes," in *Proc. IEEE CVPR*, 2017.
- [22] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, "A benchmark for the evaluation of RGB D SLAM systems," in *Proc. IEEE/RSJ IROS*, 2012, pp. 573–580.
- [23] R. T. Azuma, "A survey of augmented reality," *Presence: Teleoperators and Virtual Environments*, vol. 6, no. 4, pp. 355–385, 1997.
- [24] P. Fraga Lamas, T. M. Fernandez Carames, O. Blanco Novoa, and M. A. Vilar Montesinos, "A review on industrial augmented reality systems for the Industry 4.0 shipyard," *IEEE Access*, vol. 6, pp. 13358–13375, 2018.

- [25] E. Glaessgen and D. Stargel, "The digital twin paradigm for future NASA and U.S. Air Force vehicles," in Proc. AIAA Struct., Struct. Dyn. Mater. Conf., 2012.
- [26] C. Gehrman and M. Gunnarsson, "A digital twin based industrial automation and control system security architecture," IEEE Trans. Ind. Informat., vol. 16, no. 1, pp. 669 680, 2020.
- [27] N. Mohammadi and J. E. Taylor, "Smart city digital twins," in Proc. IEEE Symp. Ser. Comput. Intell. (SSCI), 2017, pp. 1 5.
- [28] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," Nature Machine Intelligence, vol. 1, pp. 206 215, 2019.

IJETS