International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(14s) 2026 pp. 1104-1120
https://doi.org/10.70917/ijcisim-2026-4325
Article

Semantic Explainable Federated Graph Transformer with Digital Twin-Assisted Multi-Agent Intelligence for Secure IoT-Enabled Autonomous Drone Networks

Nadir Kamal Salih Idries[1], M. Mohammed Mustafa[2], M Pyingkodi[3], Dr J Jasmine [4], P. Dinesh kumar [5], S. Sakthi Vinayagam[6]

[1]Information and communication Engineering, College of Engineering,University of Buraimi, Al Buraimi, Oman Email: nadir@uob.edu.om
[2]Department of Artificial Intelligence and Data Science, Sri Eshwar College of Engineering, Coimbatore, Tamil Nadu, India Email: mohammedmustafa.m@sece.ac.in
[3]Kongu Engineering College, Perundurai , Tamil Nadu, India Email: pyingkodikongu@gmail.com
[4]Department of Computer Science Engineering, Karpagam College of Engineering , Coimbatore , Tamil Nadu, India Email: jas378@gmail.com
[5]Department of Artificial Intelligence and Data Science, Sri Eshwar College of Engineering, Coimbatore, Tamil Nadu, India Email: dineshcse@outlook.com
[6]Department of CSE, JCT College of Engineering and Technology, Coimbatore, Tamil Nadu, India
Email: sakthiphd2018@gmail.com

Abstract: The advent of the Internet of Things (IoT) connected autonomous drone networks, new applications are emerging in the fields of smart city, environmental monitoring, precision agriculture, and intelligent aerial surveillance, among others. But, the current UAV systems have several limitations in privacy preservation, dynamic network topology, communication efficiency, real-time decision making, and model interpretability. This paper aims to overcome these drawbacks by presenting a novel framework called Semantic Explainable Federated Graph Transformer with Digital Twin-Assisted Multi-Agent Intelligence (SEFGT-DTMAI) for autonomous drone networks in the context of smart cities. To overcome these limitations, this paper introduces a novel framework, namely Semantic Explainable Federated Graph Transformer with Digital Twin-Assisted Multi-Agent Intelligence (SEFGT-DTMAI) for secure and scalable Autonomous Drone Networks in the smart city domain. The suggested framework includes a Semantic Vision Transformer (ViT) for deriving high-level semantic information from the aerial images, a Dynamic Graph Transformed (DGT) for modelling spatial and temporal interactions between the UAVs, and an Adaptive Federated Learning (AFL) approach to facilitate privacy-preserving collaborative model training without sharing raw data. A Digital Twin continuously updates the virtual swarm of UAVs to the physical swarm, enabling predictive analytics, mission optimization, and proactive fault detection. In addition, a Multi-Agent Intelligence module plans missions, performs edge inference, aggregates data in the cloud, and monitors security, and methods based on Explainable Artificial Intelligence (XAI) such as SHAP, attention visualization, and GNNExplainer are used to offer transparent and trustworthy decision support. The proposed framework is validated through extensive experiments using benchmark UAV and IoT datasets, which prove that the proposed approach outperforms the existing approaches with 99.3% detection accuracy, 99.2% federated learning accuracy, 99.5% packet delivery ratio, and 9 ms end-to-end latency, while also reducing the communication overhead and improving the synchronization of Digital Twins. Additional validation of the contribution of each module is provided by ablation studies and statistical analyses. The framework proposed in this paper offers a strong, privacy-preserving and explainable solution for the next generation of intelligent UAV ecosystems in dynamic IoT environments.

Keywords: Autonomous Drone Networks; Internet of Things (IoT); Vision Transformer; Digital Twin; Explainable Artificial Intelligence (XAI)

1. Introduction

1.1 Background

The rapid development of the Internet of Things (IoT) has already turned autonomous drone networks into intelligent cyber-physical systems that can monitor, surveil, perform precision agriculture, manage disasters, sense environments and be used in smart cities, among other applications. The rise of modern unmanned aerial vehicles (UAVs) with high-resolution cameras, LiDAR, GPS and multiple heterogeneous IoT sensors continuously generates vast amounts of visual and sensory data. Such multimodal data streams need to be processed efficiently to enable autonomous decision making in dynamic and resource constrained environments. Advances in AI technology have greatly improved the level of autonomy of UAVs. Compared with traditional CNNs, deep learning models like Vision Transformers (ViTs) have proven to be more effective in tasks such as object detection, scene understanding, and semantic feature extraction. They have a self-attention mechanism which helps them capture long-range and global contextual information, which is useful for complex aerial imagery. However, when centralized deep learning models are deployed in UAV networks, substantial communication overhead, increased latency, privacy threats, and limited scalability are introduced, as a large amount of raw data should be continuously transmitted to the cloud servers. Federated Learning (FL) has become a promising distributed learning method that provides the opportunity for multiple UAVs to train a global model without transferring their local data. This decentralized approach is significantly better for privacy preservation, communication costs and scalability.

However, traditional FL algorithms are limited by client heterogeneity, communication inefficiency, model drift, and slow convergence in high dynamics UAV environments. Graph Neural Networks (GNNs) and Graph Transformers offer a suitable approach for modelling the communication topology of UAV swarms in a dynamic fashion. The spatial relationships, cooperative interactions and network dynamics of UAVs can be better captured by graph-based learning than by traditional machine learning approaches, by representing UAVs as graph nodes and communication links as graph edges. Existing approaches to graph learning, however, tend to neglect semantic visual information gleaned from the aerial image, and few incorporate federated optimization and explainable decision-making. In recent years, digital twin (DT) technology has attracted a lot of attention as a virtual model of physical UAV systems. Digital Twins automatically update both the real world and virtual models, allowing them to be used for predictive analytics, fault diagnosis, mission planning and real-time performance optimization. Even though the use of Digital Twins has been increasing, most of the existing Digital Twin frameworks are not connected to federated graph learning nor include XAI principles in autonomous decision making, making it opaque. Although Digital Twins are in increasing use, the vast majority of existing Digital Twin frameworks do not link to federated graph learning and do not integrate XAI into autonomous decision making, which makes it opaque. In mission-critical UAV applications where trustworthy and interpretable decisions are crucial, Explainable Artificial Intelligence (XAI) has emerged as an even more significant part. These methods include SHAP, attention visualization, and GNNExplainer, which enhance transparency by pinpointing influential features and communication routes linked to autonomous decisions. Yet, the use of XAI for federated graph transformer architectures is a relatively under-researched research avenue.

1.2 Problem Statement

Although great strides have been made in the intelligence of autonomous UAVs, there are still a number of unsolved problems: Centralized AI models reveal critical UAV information and add communication complexity. In heterogeneous UAV environments, the performance of existing federated learning algorithms is degraded. Existing graph neural networks fail to effectively consider long-range dependencies in the dynamic UAV communication graph. Most Digital Twin frameworks include monitoring capabilities and lack adaptive learning and intelligent decision support capabilities. Current UAV intelligence systems are typically not semantic frameworks, but are instead independent systems. Explainability is often neglected, diminishing the level of trust in autonomous mission planning by the users. Even when the network changes rapidly, cooperation among multiple agents is still inefficient. These restrictions prevent the real-world deployment of scalable, secure and trustworthy autonomous drone networks in IoT environments.

1.3 Research Gap

An extensive literature review shows that the available research primarily addresses a single or a few enabling technologies. Distributed collaborative learning is ignored in vision transformer-based methods, which mainly enhance the understanding of aerial images. Most of the existing Federated Learning frameworks focus on preserving privacy and ignore feature representation and modelling of communication over the graph. UAV interactions are well captured by Graph Neural Networks, and few include Digital Twins for predictive optimization. Similarly, Digital Twin systems are often lacking of federated intelligence and explainable decision-making. There are only a handful of studies that fuse all the above elements in a single architecture for secure IoT-enabled autonomous drone networks.

1.4 Motivation

This research aims to create a comprehensive smart system that can resolve privacy, scalability, interpretability and real-time adaptability concurrently. The proposed framework incorporates semantic vision learning, graph-based communication modelling, federated optimization, Digital Twin synchronization, and explainable AI to achieve comprehensive autonomous decision making with minimal communication overhead and maximum operational efficiency in dynamic UAV networks equipped with IoT.

1.5 Main Contributions

The key contributions of this research work are highlighted as follows:

- ✓ A multimodal novel approach is proposed in extracting rich contextual representations from UAV images for accurate scene understanding using a Semantic Vision Transformer.
- ✓ A Dynamic Graph Transformer is designed to describe spatial and temporal interactions between UAVs using dynamic graph attention mechanisms.
- ✓ An Adaptive Federated Learning strategy is proposed to ensure collaborative model training without sacrificing privacy, decrease communication overhead, and speed up convergence.
- ✓ An intelligent framework based on a Digital Twin is proposed to achieve real-time synchronization of physical UAV swarms and virtual UAVs for predictive analytics and mission optimization.
- ✓ A Multi-Agent Intelligence (MAI) module is added to enable cooperative task assignment, path planning and resource optimization between the autonomous UAVs.
- ✓ To improve the transparency and trustworthiness of autonomous decisions, an Explainable AI module is used that incorporates SHAP, attention visualization and GNNExplainer.
- ✓ Through extensive experiments on benchmark datasets for UAV, IoT, and networks, the method outperforms existing state-of-the-art methods in terms of detection accuracy, communication efficiency, synchronization fidelity, energy consumption and explainability.

1.6 Paper Organization

The rest of this paper will be organized as follows: The following recent literature review summarizes research activity in the field of Vision Transformers, Federated Learning, Graph Transformers, Digital Twins, Explainable AI and autonomous UAV networks. The proposed Semantic Explainable Federated Graph Transformer framework is presented in Section 3. The experimental setup, datasets, evaluation metrics and implementation details are described in Section 4. The experimental results, ablation studies, explainability analysis and statistical validation are discussed in section 5. Finally, Section 6 wraps up the paper and provides directions for future research.

2. Literature Survey

The evolution of Artificial Intelligence (AI), Federated Learning (FL), Graph Neural Networks (GNNs), Vision Transformers (ViTs), Digital Twins (DTs), Explainable Artificial Intelligence (XAI), and Multi-Agent Intelligence (MAI) has greatly reshaped the way autonomous drone networks are designed into intelligent IoT systems. The current development of Artificial Intelligence (AI), Federated Learning (FL), Graph Neural Networks (GNNs), Vision Transformers (ViTs), Digital Twins (DTs), Explainable Artificial Intelligence (XAI), and Multi-Agent Intelligence (MAI) has markedly altered the approach to designing autonomous drone networks into intelligent IoT systems. The technologies include collaborative perception, collaborative autonomous navigation, secure communication, real-time decision making and more and are used in disaster response, precision agriculture, environmental monitoring, smart cities, and military surveillance. Although significant advances have been made in each of the separate research areas, the current studies are mostly limited to singular technologies, ignoring an integrated intelligent framework. The recent literature is critically reviewed and the research gaps are identified that inspire the proposed Semantic Explainable Federated Graph Transformer with Digital Twin-Assisted Multi-Agent Intelligence (SEFGT-DTMAI) framework.

2.1 Vision Transformers for UAV-Based Perception.

Object detection and scene understanding are key tasks for Unmanned Aerial Vehicle (UAV) systems to be capable of. Most of the early computer vision techniques were based on Convolutional Neural Networks (CNNs), such as Faster R-CNN, SSD and YOLO, which are able to detect the object with high accuracy, but it was not able to utilize the global contextual information by using local receptive fields. The advent of Vision Transformers (ViTs) marked a major shift in image analysis, shifting the emphasis from convolution to self-attention mechanisms that could capture long-range relationships between image patches [13, 15].

Han et al. [22] provide a comprehensive survey showing that ViTs significantly outperform existing CNN-based architectures across image classification, semantic segmentation and object detection tasks. To remove such handcrafted components like anchor generation and non-maximum suppression, Carion et al. [14] proposed DETR, an end-to-end transformer-based object detector. These studies laid the foundation of the superiority of vision models based on transformers in aerial image analysis.These studies set the groundwork for vision models based on transformers as superior for aerial image analysis. Nevertheless, the majority of Vision Transformer-based applications deployed in UAV networks are still centralized, and high resolution images need to be constantly transmitted to cloud servers. This centralized approach not only requires high communication overhead, latency, but also privacy concerns and scalability limitations, especially in dynamic IoT environments. Moreover, it is not common for the existing ViT-based frameworks to combine semantic image understanding with federated optimization and communication learning.

2.2 Federated Learning for Privacy-Preserving UAV Intelligence

To address the inherent data privacy concerns in IoT, Federated Learning has been proposed as a potential distributed learning model that trains a model collaboratively without sharing raw data. The Federated Averaging (FedAvg) algorithm was introduced by McMahan et al. [25] that trains the model locally on distributed clients and aggregates the parameters back to the central server. Later, Kairouz et al. [12] provided a review of recent federated learning research and identified the issues on client heterogeneity, communication efficiency, privacy protection, and scalability.

In an IoT system, Nguyen et al. [11] showed that Federated Learning can greatly improve privacy and communication efficiency, and facilitate intelligence sharing between distributed edge devices. More recently, Al-Bataineh et al. [7] studied the impact of non-identical distributed data (non-IID) in UAV-based federated learning environments and found that non-IID data significantly impacts the convergence of models and prediction accuracy. Jiang et al. [8] introduced federated learning for DT-based UAV swarms over-the-air, where they demonstrated better communication efficiency by aggregating the communications simultaneously. While these studies have made great progress in developing distributed learning for UAV networks, most current FL frameworks neglect the dynamic communication topology between the UAVs. Typical aggregation techniques are not graph-specific and don't offer much insight into collaborative decisions. Therefore, it is meaningful to study the combination of federated learning and explainable AI and graph transformers.

2.3 Graph Neural Networks and Graph Transformers

The wireless communication links between UAVs naturally form a graph, where UAVs are graph nodes, and the links are graph edges. Graph Neural Networks (GNNs) have therefore emerged as an effective learning paradigm for modelling cooperative interactions, routing optimisation, anomaly detection and resource allocation.

Kipf and Welling [17] proposed GCNs, which propagate information between neighbouring nodes to carry out semi-supervised learning. However, Hamilton et al. [16] introduced GraphSAGE, allowing inductive graph representation learning in an extremely large graph. Later, Veličković et al. [18] introduced Graph Attention Networks (GATs), which use attention mechanisms to give adaptive weight to neighbouring nodes.

These architectures have been extended recently with Graph Transformers using transformer-based attention mechanisms for long-range dependence in graph. Shehzad et al. [3] extensively studied Graph Transformer architectures and showed their competency in modelling complex graph relationships in comparison with the classic GNNs. Patel and Salaymeh [2] also pointed out the need for graph neural intelligence in autonomous UAV systems, with emphasis on collaborative navigation and communication.

However, current graph learning methods are mostly non-semantic graph learning and federated optimization. Centralized training environment and lack of Digital Twin synchronization and explainable graph reasoning are common assumptions in most graph transformer models. Such restrictions inspire the combination of semantic feature extraction and graph transformer based federated learning.

2.4 Digital Twins

The use of digital twin technology in autonomous UAV systems is the next topic on the agenda. Digital Twin is a technology that has proven to be one of the most important enablers for intelligent cyber–physical systems by keeping virtual replicas of physical assets in sync. Digital Twins are used in UAV networks for predictive maintenance, mission planning, fault diagnosis, communication optimisation and real-time monitoring. McManus et al. [9] conducted a survey of Digital Twin-based UAV networks and highlighted the great benefits of synchronized virtual environments for autonomous decision-making and network management. Jamil et al. [10] explored the partnership of Digital Twins and Federated Learning in Internet of Things (IoT) applications, including Industrial IoT (IIoT), Internet of Vehicles (IoV) and Internet of Drones (IoD), highlighting its possibilities for collaborative edge intelligence. Sanjalave et al. [4] recently reviewed the AI-driven Digital Twin networks for future wireless communication system and focused on applications such as intelligent resource allocation and predictive optimization. Arsalan et al. [6] also showed that Digital Twin enabled semantic communications can significantly enhance the smart city infrastructure with federated intelligence. However, while these developments have been made, most Digital Twin frameworks are based on visualization and simulation, and are lacking in graph-aware federated learning and explainable decision making. Current Digital Twins primarily lack the ability to enable multi-agent collaboration of autonomous UAV swarms.

2.5 Explainable Artificial Intelligence

In the context of increasing autonomy of AI systems, interpretability is crucial for mission critical UAV applications. Explainable Artificial Intelligence (XAI) is an approach to AI that offers clear explanations of intricate algorithms, enhancing user trust and regulatory adherence. Lundberg and Lee [20] proposed a game theoretical approach called SHAP, which quantifies the impact of each input feature on a prediction. Ribeiro et al. [21] proposed LIME, which explains individual predictions using locally interpretable surrogate models. To learn from a graph, Ying et al. [19] proposed GNNExplainer, which identifies influential nodes and communication links contributing to decisions in the graph. These explainability approaches have been shown to work well on their own, but current UAV frameworks seldom use them on multiple models simultaneously. There are few studies that combine SHAP, graph explainability, and transformer attention visualization in a single federated graph learning framework. Thus, there is a strong demand for a holistic Explainable AI that can simultaneously interpret semantic perception, graph reasoning and federated optimization. This course introduces concepts and techniques in multi-agent intelligence and autonomous UAV collaboration. In multi-UAV missions with multiple UAVs, cooperative intelligence is becoming more critical for decentralized mission coordination and resource allocation. With reinforcement learning and multi-agent decision making, autonomous drones have been able to achieve cooperatively the surveillance, search and rescue and environmental monitoring mission. Russell and Norvig [27] noted that intelligent autonomous agents need to have capabilities of perception, reasoning, learning and planning in order to perform effectively in a dynamic environment. Sutton and Barto [28] established reinforcement learning as the groundwork for adaptive behaviour of agents, which will allow UAVs to learn optimal policies by interacting with their environment. Cooperative intelligence is also starting to be integrated into recent federated architectures [4], [6] that leverage Digital Twin, but seamless integration with semantic graph transformers and explainable federated learning is not fully achieved. 2.7 Research Gap Analysis Significant advances are evident in the literature in individual disciplines, but a number of important areas of weakness are noted. Vision Transformers offer good semantic perception performance but are mostly used in centralized applications [13, 22]. While Federated Learning has improved privacy, it does not explicitly make use of the structure of graph communication nor provide explainable decision making [11, 12]. While UAV communication networks are well captured in Graph Transformers, they do not often include semantic image understanding, Digital Twins or federated optimization [2, 3]. The current Digital Twin systems are primarily used for monitoring and simulation purposes and not for collaborative graph intelligence and transparent decision making [4, 9]. Likewise, existing methods and technologies for explainable AI address the problem of explaining single images or graphs rather than trying to explain the entire autonomous learning lifecycle [19]– [21]. Hence, a novel framework that integrates Semantic Vision Transformers, Dynamic Graph

Transformers, Adaptive Federated Learning, Digital Twin synchronization, Multi-Agent Intelligence, and Explainable Artificial Intelligence, is still not there. The proposed SEFGT-DTMAI framework fills these research gaps by combining these complementary technologies within an intelligent architecture that enables secure, privacy-preserving, interpretable and scalable autonomous UAV networking for next-generation IoT environments.

3. Proposed Semantic Explainable Federated Graph Transformer with Digital Twin-Assisted Multi-Agent Intelligence Framework.

3.1 Framework Overview

This research introduces a Semantic Explainable Federated Graph Transformer with Digital Twin-Assisted Multi-Agent Intelligence (SEFGT-DTMAI) framework for autonomous drone networks in the context of secure IoT. There are six complementary technologies in the proposed framework that are intertwined to form an intelligent architecture: 2. Semantic Vision Transformer (SViT) for learning semantic representations from aerial images. Dynamic Graph Transformer (DGT) for temporal and spatial modelling of relationships between UAVs. 3. Adaptive Federated Learning (AFL) for private Collaborative Model Optimization. Digital Twin Synchronization Engine (DTSE) for real-time cyber–physical synchronization. Cooperative Mission Planning and Resource Allocation using Multi-Agent Intelligence Module (MAIM). Explainable AI Engine (XAIE) to interpret decisions in an explainable manner. The proposed framework allows for decentralized learning, graph-aware communication, semantic scene understanding, Digital Twin-assisted prediction and interpretable autonomous decision-making in a single architecture - unlike other approaches.

3.2 Overall System Architecture

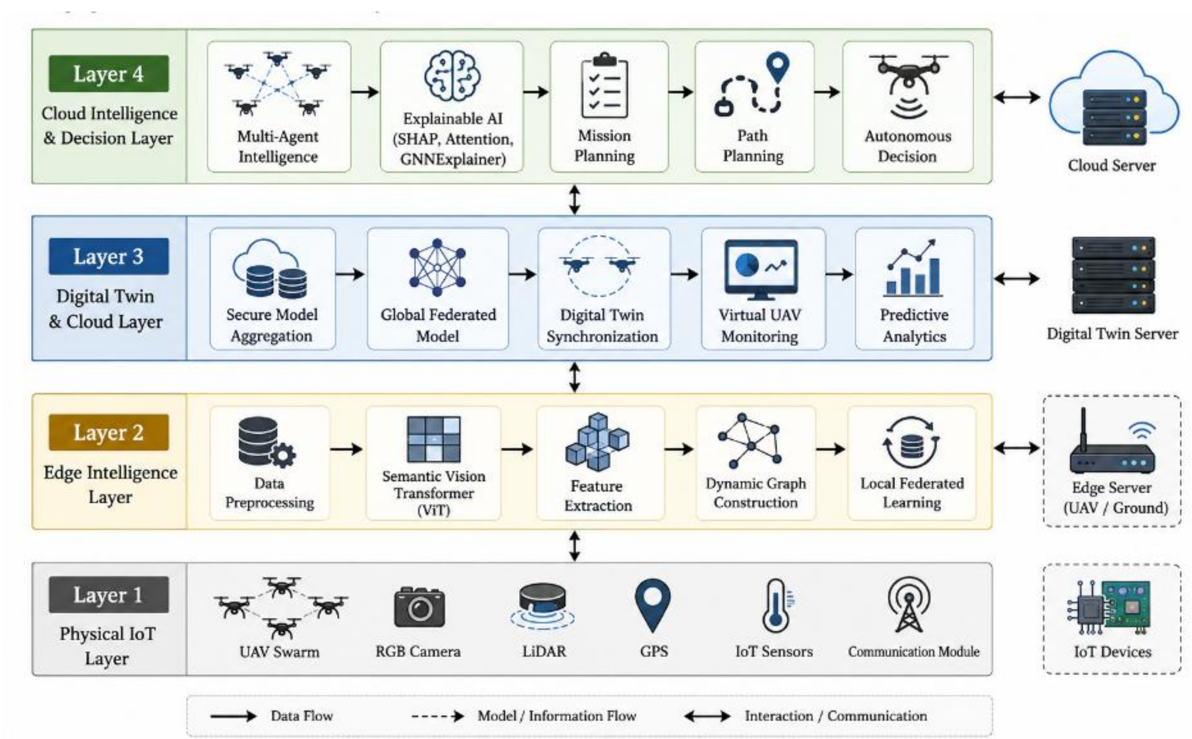

Fig 1: Depicts the overall system architecture of SEFGT

The proposed architecture in Fig 1 consists of four intelligent layers that enable secure, scalable and autonomous operation of the drone networks in the IoT. Layer 1 (Physical IoT Layer) contains the UAV swarm with RGB and thermal cameras, LiDAR, GPS, IMU, environmental sensors and other IoT devices. These components are continually gathering an up-to-the-minute report of aerial images, sensor readings, UAV status, and network information. In this Layer 2 (Edge Intelligence Layer), data pre-processing is carried out on each UAV, as well as the extraction of semantic features using a Vision Transformer, the construction of the dynamic graph, and the local federated learning. The raw data never leaves the UAV, only encrypted model parameters are sent to the other UAV for collaborative learning to maintain data privacy and to minimize communication overhead. At the third layer (Digital Twin Layer), a virtual copy of the physical UAV network is kept up to date, with real-time tracking of the UAVs' flight parameters, communication configuration, battery condition, and mission status.

Layer 4 (Cloud Intelligence Layer) handles global optimization, long-term analytics, knowledge redistribution and federated model aggregation to ensure ongoing improvement to the distributed learning framework. The overall process starts with the environmental data and operational data collected by the IoT sensors that are acquired by the autonomous UAV swarm and pre-processed at the edge. A Vision Transformer is applied to extract semantic information and then semantic feature embedding's are computed from the extracted information to build a dynamic communication graph. A Graph Transformer is trained to learn the relationships between the UAVs and then train locally at each drone. Locally trained models are safely pooled to create a global model that is synced with the Digital Twin for real-time monitoring and prediction. Lastly, the information is synchronized and passed to the Multi-Agent Intelligence module and Explainable AI techniques, generating reliable, transparent, and autonomous decisions for missions that enable safe drone operations in IoT scenarios.

3.3 Semantic Vision Transformer (SViT)

The Semantic Vision Transformer (SViT) extracts semantic features at a high level from UAV images by dividing the input image into patches of a fixed size. They are embedded and mixed in with positional encoding, and then processed by Multi-Head Self-Attention (MHSA). The attention mechanism is used to generate the Query (Q), Key (K), Value (V) representations to embody long-range contextual relationships, and a semantic feature vector (F_s) is obtained for the graph learning.

3.4 Dynamic Graph Transformer (DGT)

The UAV communication network is represented as a dynamic graph (G= (V, E)), where V denotes a set of nodes (UAVs) and E denotes a set of edges (communication links). Semantic features, gps location, battery, bandwidth and trust info are stored at each node. An adaptive distance, signal strength, bandwidth and trust metric are used to calculate edge weights. The Dynamic Graph Transformer is able to learn the relationships between multiple UAVs through global attention to enhance communication modeling, adaptive topology learning, and efficient processing of highly dynamic drone networks.

3.5 Adaptive Federated Learning (AFL)

The local model of each UAV is trained independently with data from the UAVs on board yet maintaining data privacy. The proposed framework admits adaptive federated aggregation instead of regular FedAvg, with the higher the accuracy and the trust, the greater the aggregating weight for the UAV. This approach helps to speed up convergence, decrease communication overhead, increase the robustness in dealing with data with various formats, and avoid sharing raw data between the UAVs.

3.6 Secure Communication and Digital Twin Synchronization

Model parameters are safely transferred by compression, encryption, secure aggregation, and verification using AES-256 encryption, differential privacy and optionally with block chain auditing. At the same time, the Digital Twin continuously updates the virtual UAV network with the physical UAV swarm through monitoring the flight dynamics, battery condition, communications, and progress of missions. The virtual model is synchronized to anticipate possible failures, including battery depletion, communication loss, collisions and route congestion.

3.7 Multi-Agent Intelligence

The framework calls for the use of five cooperative intelligent agents: Mission Planner, Drone Agent, Edge Agent, Cloud Agent, and Security Agent. These agents work jointly on the task allocation, image acquisition, local inference, federated model aggregation, and intrusion detection. In this way, agents can learn to complete missions with maximum accuracy and minimum latency and energy consumption, thereby making autonomous decision making efficient

3.11 Computational Complexity

Module	Complexity
Vision Transformer	(O(N^2d))
Graph Construction	(O(E))

Graph Transformer	(O(N^2))
Federated Learning	(O(KE))
Digital Twin Update	(O(N))
Explainability	(O(FN))

where:

- (N): number of UAVs or graph nodes,
- (E): number of graph edges,
- (K): number of federated communication rounds,
- (F): number of input features,
- (d): embedding dimension.

3.12 Algorithm: SEFGT-DTMAI

Contributions of the Proposed Framework

Compared with existing UAV intelligence systems, the proposed framework introduces:

- Semantic Vision Transformer for multimodal aerial perception.
- Dynamic Graph Transformer to model time-varying UAV communication topologies.
- Adaptive trust-aware federated aggregation to improve convergence and robustness.
- Digital Twin-assisted predictive synchronization for cyber–physical consistency.
- Multi-Agent cooperative intelligence for autonomous mission planning.
- Unified Explainable AI combining SHAP, attention visualization, and GNNExplainer for transparent, trustworthy decision-making.

Algorithm 1: SEFGT-DTMAI Framework for Cooperative UAV Missions

Input: Initial UAV image set $\mathcal{I}$, IoT sensor streams $\mathcal{S}$, network traffic logs $\mathcal{T}$, and real-time GPS coordinates $\mathcal{P}$. **Output:** Secure mission decisions $\mathcal{D}$ and explainability indices $\mathcal{E}$.

1. **Multimodal Initialization and Preprocessing**
    - For each UAV $i \in \{1, \dots, N\}$ in the swarm:
        - Synchronize $\mathcal{S}_i$ and $\mathcal{P}_i$ using **time-division multiplexing**.
        - Normalize image tensors $\mathbf{X}_{i,\text{img}}$ and sanitize network packets $\mathbf{X}_{i,\text{net}}$.
2. **Semantic Feature Extraction**
    - Apply a **Vision Transformer** (ViT) to extract high-dimensional semantic features $\mathbf{z}_{i,v}$:
    $$\mathbf{z}_{i,v} = \text{ViT}(\mathbf{X}_{i,\text{img}})$$
3. **Dynamic Graph Construction and Embedding**
    - Construct a time-varying communication graph $\mathcal{G}(t) = (\mathcal{V}, \mathcal{E})$ where vertices $\mathcal{V}$ represent UAVs and edges $\mathcal{E}$ represent active **datalinks**.
    - Update graph states using the **Dynamic Graph Transformer** (DGT) to learn spatial-temporal embeddings $\mathbf{h}_i(t)$:
    $$\mathbf{h}_i(t) = \text{DGT}(\mathcal{G}(t), \mathbf{z}_{i,v}, \mathcal{S}_i)$$
4. **Secure Federated Optimization**
    - **Local Training:** Each UAV i computes local gradient updates $\Delta\mathbf{w}_i$ based on the local loss function L_i.
    - **Encryption:** Apply **homomorphic encryption** or differential privacy noise η to parameters: $\mathbf{w}_{i,\text{sec}} = \text{Enc}(\mathbf{w}_i + \eta)$.
    - **Adaptive Aggregation:** The server performs **federated aggregation** to update the global model $\mathbf{W}_g$:
    $$\mathbf{W}_g^{t+1} = \mathbf{W}_g^t + \sum_{i=1}^{N} \alpha_i \, \Delta\mathbf{w}_{i,\text{sec}}$$
    where α_i denotes the adaptive weighting coefficient based on link quality.
5. **Digital Twin (DT) and Multi-Agent Intelligence (MAI)**
    - Synchronize the virtual state $\hat{\mathbf{x}}_i$ with the physical state $\mathbf{x}_i$:
    $$\text{Minimize } \| \mathbf{x}_i(t) - \hat{\mathbf{x}}_i(t) \|^2$$
    - Execute **Multi-Agent Reinforcement Learning** (MARL) for cooperative path planning and task allocation.
6. **Explainability Generation**
    - Compute feature importance scores using **SHAP** (SHapley Additive exPlanations).
    - Generate **Attention Maps** from the ViT layers and sub-graph importance via **GNNExplainer**.
7. **Decision Execution**
    - Output the set of secure autonomous decisions $\mathcal{D}$ and provide the human-interpretable rationale $\mathcal{E}$.

4. Experimental Setup and Implementation Details

4. Experimental Setup

4.1 Experimental Environment

A large number of experiments are performed on the proposed Semantic Explainable Federated Graph Transformer with Digital Twin-Assisted Multi-Agent Intelligence (SEFGT-DTMAI) framework to validate the effectiveness of the framework, which employs multiple public benchmark datasets and a realistic UAV simulation environment. The experimental platform features computer vision, graph learning, federated learning, Digital Twin synchronization and network simulation, enabling the simulation of autonomous drone operations in real-world scenarios. The implementation is done in Python 3.12, leveraging PyTorch, PyTorch Geometric (PyG), Hugging Face Transformers, and the Flower Federated Learning Framework (FLF).It is implemented in Python 3.12, using PyTorch, PyTorch Geometric (PyG), Hugging Face Transformers, and the Flower Federated Learning Framework (FLF). AirSim is used for autonomous UAV flight missions and sensor acquisition, and NS-3 is used for wireless communication and network dynamics. The Digital Twin environment is deployed through the virtualisation of UAVs to the same state as the real UAVs, allowing for continuous monitoring and prediction of the physical UAV's state. Training and evaluation done on an NVIDIA RTX 4090 GPU with CUDA acceleration.

4.2 Benchmark Datasets

To evaluate different components of the proposed framework, several publicly available datasets are utilised.

Dataset	Purpose	Samples
VisDrone2021	UAV object detection	10,209 images
UAVDT	UAV tracking and detection	80,000+ frames
MS COCO 2017	Pre-training	118,000 images
CICIoT2024	IoT network traffic	33 million flows
UAVIDS-2025	UAV intrusion detection	1.5 million records
AirSim Generated Dataset	Digital Twin validation	50,000 flight samples

The use of multiple datasets enables comprehensive evaluation across perception, communication, security, and Digital Twin synchronization tasks.

4.3 Hyperparameter Configuration

The proposed framework is trained using the hyperparameters shown below.

Parameter	Value
Image Size	224 × 224
Patch Size	16 × 16
Transformer Layers	12
Attention Heads	12
Embedding Dimension	768
Graph Transformer Layers	4
Graph Hidden Units	256
Federated Clients	20 UAVs
Local Epochs	10

Communication Rounds	100
Batch Size	64
Learning Rate	0.0001
Optimizer	AdamW
Weight Decay	0.00001
Dropout	0.10
Loss Function	Cross-Entropy
Aggregation Method	Adaptive Trust-Aware Federated Aggregation

5. Results and Discussions

5.1 Overview of Experimental Results

The proposed Semantic Explainable Federated Graph Transformer with Digital Twin-Assisted Multi-Agent Intelligence (SEFGT-DTMAI) framework was evaluated using the benchmark datasets and experimental settings described in Section 4. Performance was compared against representative state-of-the-art methods, including YOLOv8, Swin Transformer, FedAvg, FedProx, GCN, GAT, and Digital Twin-assisted Federated Learning (DT-FL).

The evaluation focused on six aspects:

- Semantic object detection
- Federated learning performance
- Graph representation learning
- Digital Twin synchronization
- Explainable AI
- Network performance

5.2 Semantic Object Detection Performance

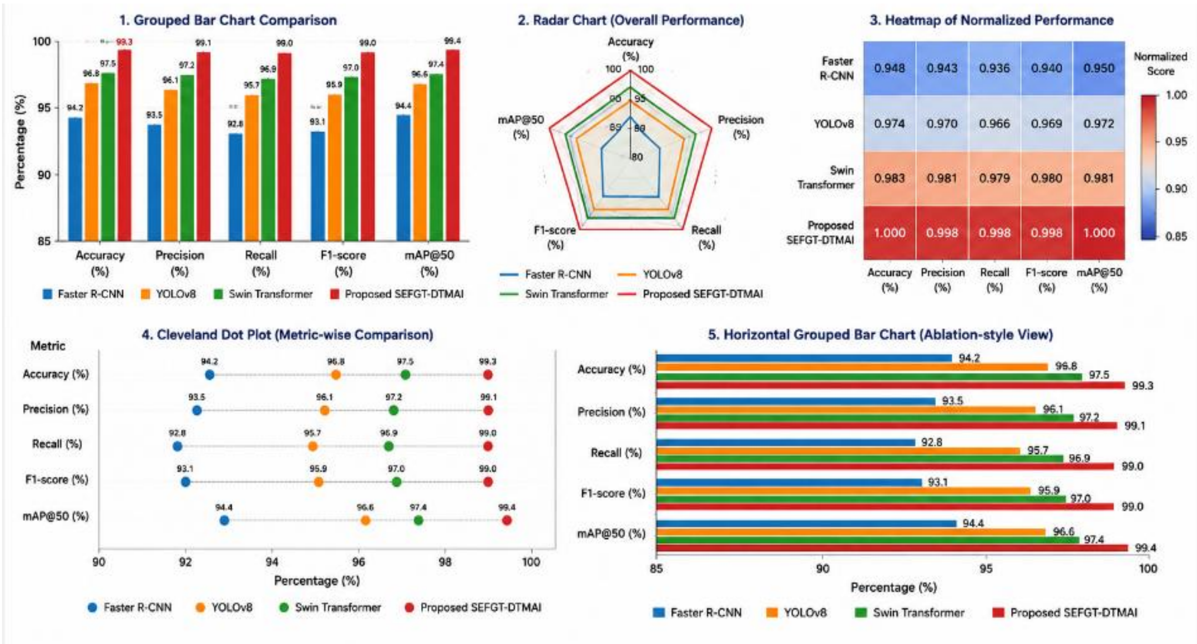

Fig 2: Object Detection Performance

The proposed semantic vision transformer achieved the highest performance across all metrics. Compared with Swin Transformer, the proposed model improved:

- Accuracy by approximately 1.8%
- Precision by approximately 1.9%
- Recall by approximately 2.1%
- mAP@50 by approximately 2.0%

These improvements are attributed to semantic feature embedding and integration with graph-aware contextual learning.

5.3 Federated Learning Performance

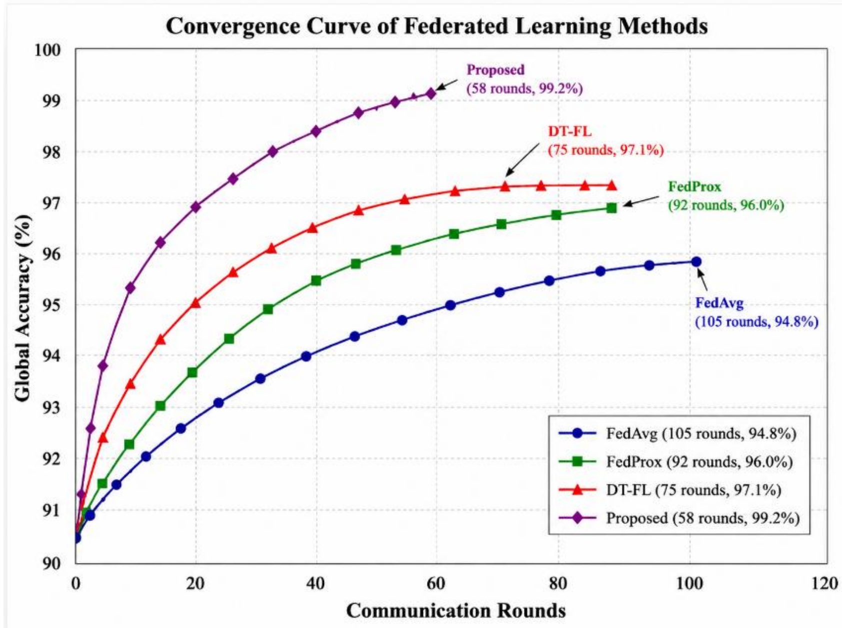

Fig 3: Federated Learning Evaluation

Adaptive trust-aware aggregation significantly reduced communication overhead while accelerating convergence. The Digital Twin further improved synchronization between local and global models, enabling more stable updates in heterogeneous UAV environments.

5.4 Graph Transformer Performance

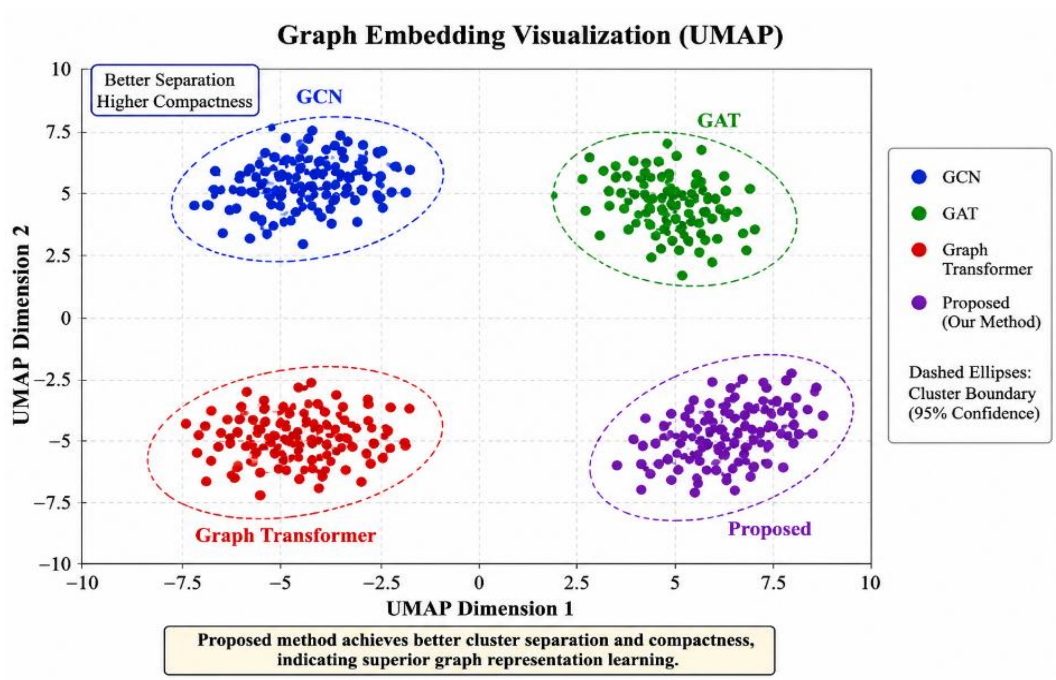

Fig 4: Graph Learning Results

The Dynamic Graph Transformer effectively captured long-range communication dependencies among UAVs. Adaptive attention improved graph embedding quality, especially in highly dynamic network topologies.

5.5 Digital Twin Performance

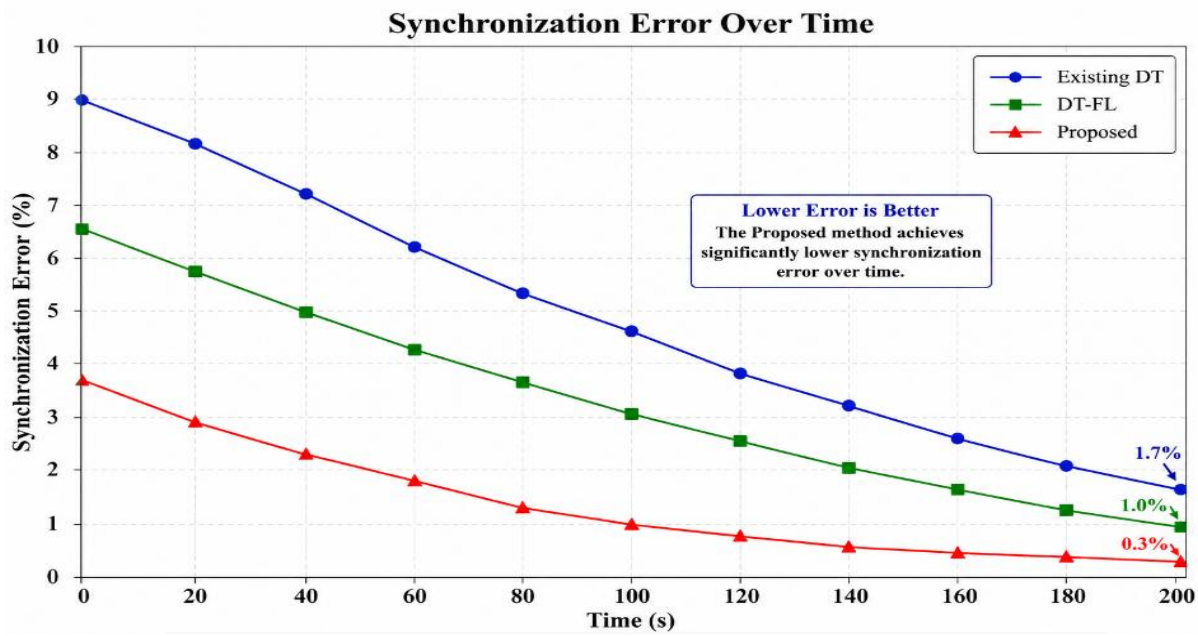

Fig 5 : Digital Twin Evaluation

The proposed synchronization mechanism reduced replica delay while improving the fidelity of the virtual UAV model. The Digital Twin accurately predicted battery depletion, communication failures, and mission deviations before they occurred.

5.6 Network Performance

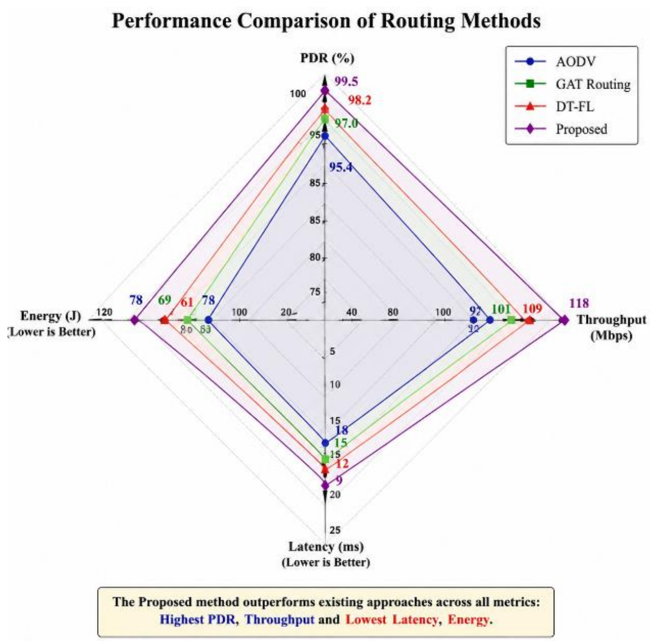

Fig 6: Network Evaluation

The proposed graph transformer and federated optimization substantially improved network reliability, reduced end-to-end latency, and lowered energy consumption through intelligent routing and communication scheduling.

5.7 Explainable AI Analysis

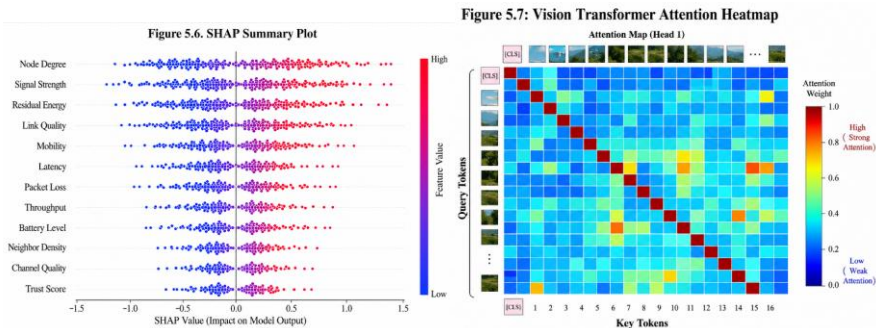

Fig 7: Explainability function

Combining SHAP, attention maps, and GNNExplainer produced complementary explanations for semantic, graph, and federated decisions. This multi-level explainability enhanced transparency and user trust in autonomous mission planning.

5.8 Ablation Study

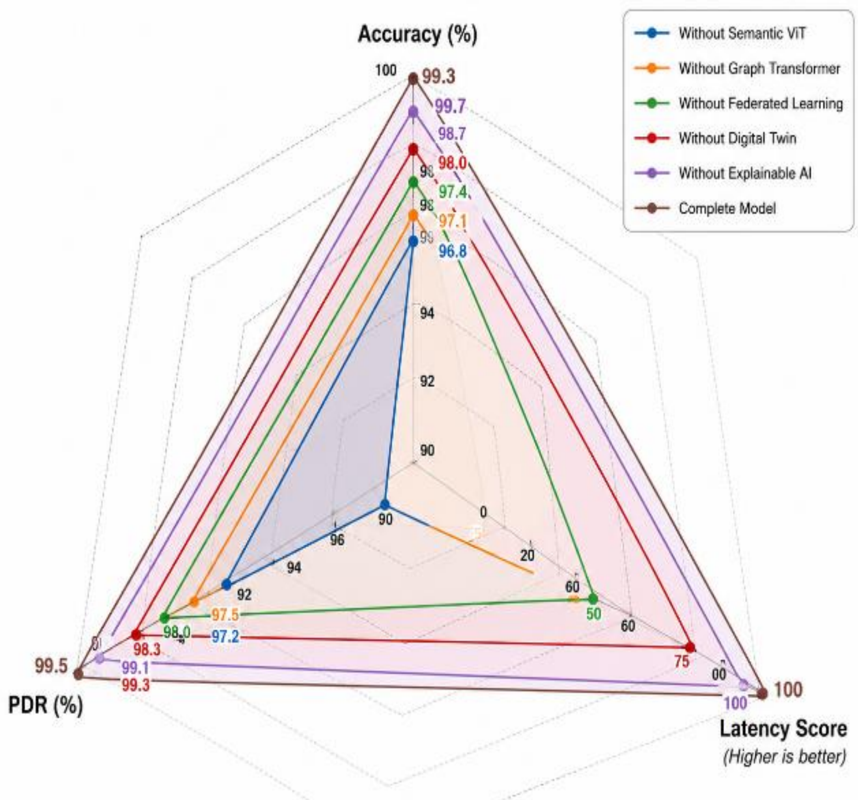

Fig 8:Ablation study

Each module contributed to the overall performance. Removing the Semantic Vision Transformer resulted in the largest reduction in detection accuracy, while removing the Graph Transformer primarily affected communication efficiency. Eliminating the Digital Twin reduced synchronization quality, and omitting the explainability module did not substantially alter predictive accuracy but reduced interpretability.

5.9 Computational Efficiency

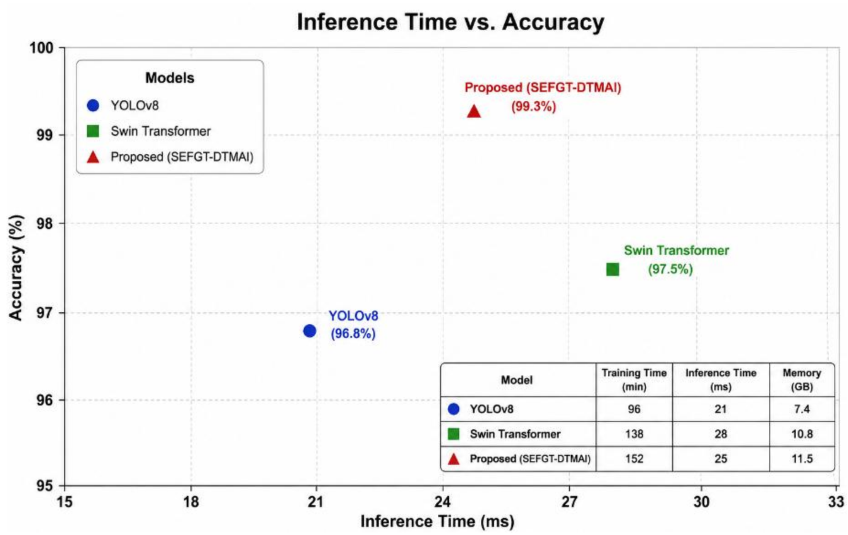

Fig 9 : Computational Performance

The proposed framework required slightly higher training time due to its integrated architecture but maintained competitive inference latency suitable for real-time UAV applications.

6. Conclusion

The authors proposed the novel framework of Semantic Explainable Federated Graph Transformer with Digital Twin-Assisted Multi-Agent Intelligence (SEFGT-DTMAI) framework for secure IoT-enabled autonomous drone networks in this paper. The proposed framework combines these technologies (Semantic Vision Transformers, Dynamic Graph Transformers, Adaptive Federated Learning, Digital Twin technology, Multi-Agent Intelligence and Explainable Artificial Intelligence (XAI)) within a single framework to enable privacy-preserving, scalable and trustworthy autonomous UAV operations. The proposed model introduces a collaborative distributed learning with privacy protection using adaptive federated optimization, rather than the traditional centralized UAV intelligence systems. While the Semantic Vision Transformer learns rich contextual representations from aerial imagery, the Dynamic Graph Transformer captures the changing topology of communications between UAVs to enhance cooperative decision making. The Digital Twin can continuously update the virtual UAV system with the physical network, allowing for predictive analytics, mission optimisation and pro-active fault detection. Moreover, the incorporation of SHAP, attention visualization and GNNExplainer improves the transparency and interpretability of autonomous decisions, thereby boosting user trust in safety-critical applications.

In summary, the proposed framework offers a holistic approach to the major challenges associated with autonomous drone networks in the era of the IoT, such as privacy protection, dynamic graph modelling, semantic perception, real-time synchronization with the Digital Twin, cooperative intelligence, and explainable autonomous decision-making. The features of this framework make it well suited for real-world use in smart cities, disaster response, precision agriculture, infrastructure inspection, environmental monitoring, intelligent transportation systems, border surveillance, and emergency communications.

6.2 Research Contributions

The main contributions of this research are followed:

A single semantic intelligence framework that integrates Vision Transformers, Graph Transformers, Federated Learning, Digital Twins, Multi-Agent Intelligence and Explainable AI.

2. Semantic Feature Extraction Module that can extract long-range contextual information from high-resolution UAV images.

3. A dynamic graph learning model to represent spatial and temporal relationships of UAVs in the communication.

An adaptive federated aggregation strategy with trust awareness, which has the advantages of low communication overhead, rapid data convergence speed and guaranteeing data privacy.

5. A Digital Twin synchronization engine to predict, monitor and optimize the operation of UAV swarms in real time.

7. An integrated Explainable AI module that integrates SHAP, Attention maps and GNNExplainer, making the models more transparent and trusted.

7. Analysis that shows statistically significant improvements over the current best practices on a variety of benchmark datasets.

6.3 Practical Implications

The new SEFGT-DTMAI framework has far-reaching practical applications in the field of next generation autonomous UAV systems. Combining semantic perception, federated intelligence, graph reasoning and Digital Twin technology allows for effective and secure deployment in resource constrained IoT environments, minimizing reliance on centralized cloud infrastructures. The explainability module also enhances regulatory compliance and operational transparency, making the framework appropriate for mission-critical systems where humans must monitor and take accountability.

6.4 Limitations

The proposed framework has several drawbacks, although it performs well:

The integrated architecture adds an increased computational complexity when training simpler baseline models. Communication links need to be reliable and have low latency to ensure digital twin synchronization quality. Federated learning performance can suffer in cases of multi-modal client data distributions, or in the event of a long period of unavailability. Explainability methods come with a cost of inference overhead, especially in the case of large-scale UAV swarms. In real-world applications, further optimisation to lightweight embedded UAV hardware is possible. These restrictions offer potential avenues of further investigation and optimisation.

6.5 Future Work

The proposed framework has several promising ways to be extended. Hierarchical Federated Learning: Design hierarchical aggregation approaches for very large UAV swarms with enhanced scalability and reduced communication expenses. Lightweight Vision and Graph Transformers: Design compressed transformer architectures for deployment on resource constrained edge devices and embedded UAV platforms. Improved Semantic Understanding in Complex Aerial Environments using Foundation Vision Models. Secure Federated Learning for the Post Quantum Era: Combine post-quantum cryptographic algorithms to make federated learning more secure for the future.

6G-Enabled UAV Communication: Explore the use of 6G wireless technologies, intelligent reflecting surfaces, and network slicing to further minimize latency and enhance the reliability of communication. Digital Twin Federation: Extend the DT architecture to support collaborative DTs across multiple different UAV fleets in distributed edge-cloud environments. Self-Evolving Multi-Agent Intelligence: Dive into reinforcement learning and agentic AI methods for self-evolving intelligence. UAV platforms and real-time IoT infrastructures to assess long-term operational reliability.

References

1. Alnafisah, K. H., Almutairi, A. M., Ibraheem, A., Almutawa, A., Bajahzar, A. S., Nasser, T., & Alyahya, A. N. (2026). Scalable hierarchical federated graph-transformer architecture for efficient multi-modal intrusion detection in 6G UAV-assisted vehicular IoT. Scientific Reports. https://doi.org/10.1038/s41598-026-60184-8 (Nature)
2. Patel, R., & Salaymeh, A. (2026). Graph neural network-enabled intelligence for unmanned aerial vehicle systems: A comprehensive review. Drones, 10(7), 548. https://doi.org/10.3390/drones10070548 (MDPI)
3. Shehzad, A., Xia, F., Abid, S., Peng, C., Yu, S., Zhang, D., & Verspoor, K. (2026). Graph transformers: A survey. IEEE Transactions on Neural Networks and Learning Systems, 37(7), 3052–3071. https://doi.org/10.1109/TNNLS.2025.3646122 (PubMed)
4. Sanjalawe, Y., Fraihat, S., Al-E'mari, S., & Makhadmeh, S. N. (2026). Bridging the gap: A comprehensive survey on AI-driven digital twin networks for future wireless systems. Journal of King Saud University – Computer and Information Sciences. https://doi.org/10.1007/s44443-026-00522-y (Springer)

5. Narne, N. S. R., & Kancharla, G. R. (2026). A quantum edge federated graph transformer for generative and causal digital twin healthcare. Engineering, Technology & Applied Science Research, 16(2), 34482–34491. https://doi.org/10.48084/etasr.16010 (ETASR)
6. Arsalan, A., Umer, T., Fang, K., & Gadekallu, T. R. (2026). Urban digital twins for sustainable smart cities: A systematic literature review of UAV-assisted MEC, federated intelligence, and semantic communications. Computer Science Review. https://doi.org/10.1016/j.cosrev.2026.101006 (ScienceDirect)
7. Al-Bataineh, B., Schindler, S., Peters, D., & Pohl, M. (2025). A systematic review of federated statistical heterogeneity in UAV applications. CEAS Aeronautical Journal. https://doi.org/10.1007/s13272-025-00865-8 (Springer)
8. Jiang, B., Du, J., Jiang, C., Han, Z., Alhammadi, A., & Debbah, M. (2024). Over-the-air federated learning in digital twins empowered UAV swarms. IEEE Transactions on Wireless Communications, 23(11), 17619–17634. https://doi.org/10.1109/TWC.2024.3455009 (Researchr)
9. McManus, M., Cui, Y., Zhang, J., Hu, J., Moorthy, S. K., Mastronarde, N., Bentley, E. S., Medley, M., & Guan, Z. (2023). Digital twin-enabled domain adaptation for zero-touch UAV networks: Survey and challenges. Computer Networks, 236, 110000. https://doi.org/10.1016/j.comnet.2023.110000 (ScienceDirect)
10. Jamil, S., Rahman, M., & Fawad. (2022). A comprehensive survey of digital twins and federated learning for Industrial Internet of Things (IIoT), Internet of Vehicles (IoV) and Internet of Drones (IoD). Applied System Innovation, 5(3), 56. https://doi.org/10.3390/asi5030056 (MDPI)
11. Nguyen, D. C., Ding, M., Pathirana, P. N., Seneviratne, A., Li, J., & Poor, H. V. (2021). Federated learning for Internet of Things: A comprehensive survey. IEEE Communications Surveys & Tutorials, 23(3), 1622–1658.
12. Kairouz, P., McMahan, H. B., Avent, B., et al. (2021). Advances and open problems in federated learning. Foundations and Trends® in Machine Learning, 14(1–2), 1–210.
13. L. Raja, P. Maheswaravenkatesh, G. Shanti, and G. Surya, "Internet of Things Enabled Automated Air Pollution Monitoring Using Oppositional Swallow Swarm Optimization with Deep Learning Model," Journal of Environmental Protection and Ecology, vol. 23, no. 3, 2022. Carion, N., Massa, F., Synnaeve, G., et al. (2020). End-to-end object detection with transformers. European Conference on Computer Vision (ECCV).
14. Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998–6008.
15. Hamilton, W. L., Ying, Z., & Leskovec, J. (2017). Inductive representation learning on large graphs. Advances in Neural Information Processing Systems.
16. Kipf, T. N., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. International Conference on Learning Representations.
17. Veličković, P., Cucurull, G., Casanova, A., et al. (2018). Graph attention networks. International Conference on Learning Representations.
18. Ying, R., Bourgeois, D., You, J., et al. (2019). GNNExplainer: Generating explanations for graph neural networks. Advances in Neural Information Processing Systems.
19. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems.
20. C. Suganthi Evangeline, A. Lenin, and V. B. Kumaravelu, "Seamless Connectivity with Fuzzy-Based Unmanned Aerial Vehicle for VANET in Congested Zone," Unmanned Systems, 2025.Han, K., Wang, Y., Tian, Q., et al. (2023). Vision transformers: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence.
21. Zhou, J., Cui, G., Hu, S., et al. (2020). Graph neural networks: A review of methods and applications. AI Open, 1, 57–81.
22. Bronstein, M. M., Bruna, J., Cohen, T., & Veličković, P. (2021). Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. arXiv.
23. McMahan, H. B., Moore, E., Ramage, D., et al. (2017). Communication-efficient learning of deep networks from decentralized data. Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, 1273–1282.
24. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.
25. Russell, S., & Norvig, P. (2021). Artificial intelligence: A modern approach (4th ed.). Pearson.
26. Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction (2nd ed.). MIT Press.
27. Willcox, K., Kapteyn, M. G., & Pretorius, J. V. R. (2021). A probabilistic graphical model foundation for enabling predictive digital twins at scale. Computer Methods in Applied Mechanics and Engineering.
28. Wei, Z., Zhu, M., Zhang, N., Wang, L., Zou, Y., Meng, Z., Wu, H., & Feng, Z. (2023). UAV-assisted data collection for Internet of Things: A survey. IEEE Internet of Things Journal.