Received: April 10, 2024. Revised: May 26, 2024.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Anomaly Detection Based on Hierarchical Federated Learning with Edge-Enabled Object Detection for Surveillance Systems in Industry 4.0 Scenario

Ibrahim Ali Alnajjar[1]* Laiali Almazaydeh[2] Ali Abu Odeh[3] Anas A. Salameh[4]
Khalid Alqarni[5] Anas Ahmad Ban Atta[6]

[1]*Department of Computer Science, College of Computer Information Technology (CCIT),*
American University in the Emirates (AUE), Dubai 503000, United Arab Emirates
[2]*Department of Software Engineering, Faculty of IT. Al-Hussein Bin Talal University,*
Ma'an, Jordan.College of Computer Information Technology, American University in the Emirates, Dubai 503000, United Arab Emirates
[3]*Academic Support Department, Abu Dhabi Polytechnic, P.O. Box 111499, Al Ain, United Arab Emirates*
[4]*Department of Management Information Systems, College of Business Administration,*
Prince Sattam bin Abdulaziz University, 165 Al-Kharj 11942, Saudi Arabia
[5]*Management Information Systems Department, Faculty of Economics and Administration*
King Abdulaziz University, Saudi Arabia
[6]*Department of Accounting and Finance, Middle East University, Amman, Jordan*
* Corresponding author's Email: ibrahim.alnajjar@aue.ae

Abstract: Anomaly detection from a video surveillance camera is a time-critical application that combines the capabilities of computer vision-based object detection algorithms to monitor and analyze various anomalous activities in Industry 4.0 scenarios. An intelligent video surveillance system for automated monitoring and analysis of video streams without human supervision is paramount in industrial scenarios. Nevertheless, the detection of anomalous objects is often hindered by the scarcity of data and privacy restrictions inherent in the centralized storage and processing of surveillance videos. To overcome these shortcomings, Federated Learning (FL) has emerged as a promising solution for the privacy-preserved processing of decentralized data. Despite significant advancements in computer vision, accurately identifying surveillance anomalies through object detection within resource-constrained edge networks remains a formidable challenge for FL-assisted anomaly detection. This difficulty arises from the limited computational capabilities and constrained resources inherent to edge devices, which impedes the performance and accuracy of anomaly detection algorithms relying on the object recognition method. Thus, this work proposes a hierarchical FL-assisted surveillance anomaly detection by integrating the YOLOv8n and Flownet models for motion-aware, accurate anomalous object detection. To design time-efficient anomaly detection for time-critical applications, the proposed approach applies the hierarchical FL that comprises multiple edge aggregators instead of cloud aggregators. The primary objective of adopting the hierarchical FL is to alleviate the communication costs associated with object detection tasks. By distributing the aggregation process across multiple edge nodes, the proposed approach enhances the efficiency of anomaly detection while minimizing latency, thereby ensuring timely responses. Finally, the FL-assisted detection system accurately identifies anomalous human activities in the manufacturing industry through the hierarchical aggregation associated with the local model of YOLOv8n and Flownet-based object detection in the edge network. Thus, the experimental results prove its anomaly detection ability in the surveillance vides by yielding 88.95% accuracy and 0.85 as the average Anomaly score while testing on the Avenue dataset.

Keywords: Anomaly detection, Hierarchical federated learning, Industry 4.0, Object detection, Surveillance video, And manufacturing industry.

1. Introduction

With the emergence of Industry 4.0, manufacturing processes have witnessed a dramatic technological change characterized by historically unprecedented levels of automation. Tiny and intelligent smart sensing equipment with cutting-edge capabilities is the foundation of Industry 4.0, bringing in a new era of interconnected manufacturing systems. The resource-constrained devices, which are also known as cyber-physical systems, function with constrained energy, memory, and bandwidth [1]. However, they are essential for accurately and quickly coordinating intricate manufacturing processes. Object detection-based video surveillance plays a pivotal role in assuring the security, safety, and efficiency of the industrial 4.0 atmosphere [2]. Anomalous object detection is vital for industrial security, achieved through video surveillance analysis. While YOLO variants excel at object detection, detecting anomalies poses challenges. To address this, pretraining YOLO models with anomalous behaviors or fine-tuning for anomaly detection is crucial. Motion feature extraction across sequential frames enhances understanding of human activities, aiding in anomaly detection alongside traditional deep learning models.

In the context of Industry 4.0, the proliferation of smart devices results in the generation of vast quantities of data, often stored in cloud-based databases. These repositories serve as the primary storage solution for the copious amounts of information produced by Industry 4.0 devices. To execute sophisticated real-time decision-making in industrial applications, access to data from these cloud servers is necessary. Minimal delay has a significant impact on the efficiency of the industry, which emphasizes the importance of responding quickly and intelligently to new situations or anomalies in industrial settings. By bringing cloud services closer to the devices, edge computing reduces the latency of the object detection model [3], [4].

Due to a multitude of privacy considerations, there exists a lack of comprehensive understanding regarding abnormal events across various Industry 4.0 environments, consequently impeding the efficacy of automated object detection models. Additionally, the substantial expense associated with transmitting large surveillance videos to centralized storage facilities serves as a barrier, constraining the real-time prediction capabilities of these models. To counter privacy concerns and high-cost issues, Federated Learning (FL) is referred to as a promising solution, as it can provide distributed model training among different industries in a privacy-preserving way [5, 6]. The FL introduces dynamic learning strategies in which the globally shared models can be updated periodically and enhanced according to the evolving conditions in diverse parts of the various industrial 4.0 environments [7, 8]. Thus, it is highly adaptive to address the unique characteristics of local industrial environments effectively. It maximizes the accuracy and robustness of FL-enabled object detection-based surveillance in distinct industrial scenarios [9]. Despite its advantages, the FL algorithm introduces significant communication overhead between edge devices and centralized cloud servers during model updates and aggregation phases, particularly in the context of cybersecurity. The high communication burden has the potential to impact system efficiency, as it leads to delays in model convergence and increased resource consumption, affecting the overall performance of the FL-based cybersecurity framework [10]. Hence, the FL algorithm needs continuous model updates and multiple aggregators to enhance the accuracy of the global model due to the data distributions in various organizations are different and demand minimal communication costs in a resource-constrained environment. Due to the resource-intensive nature of FL, deploying it with the edge environments is essential, necessitating the development of precise object detection strategies to enhance the performance of anomalous activity detection. In accordance with this, the edge-assisted federated approach ensures that computational resources are efficiently utilized at the edge, enabling effective anomaly detection while minimizing the burden on the underlying infrastructure. As FL consumes high resources, running it on an edge environment and designing precise object detection strategies to improve the performance of abnormal activity detection is crucial. Thus, this work designs the hierarchical FL model associated with anomaly detection in industry by examining videos.

1.1 Contributions

This work aims to propose an FL-enabled human anomaly detection model for an Industry 4.0 surveillance system in which different human abnormalities are recognized in video sequences with minimal communication cost and higher accuracy in a privacy-preserving manner. The major contributions of the proposed model are as follows.

- The primary aim of the proposed model is to design a surveillance anomaly detection system for Industry 4.0 with the assistance of a hierarchical FL and YOLOv8 object detection model. It includes three major steps: data collection with feature extraction, FL-assisted edge computing, and YOLOv8-based object detection.

- The proposed anomaly detection model utilizes a hierarchical FL architecture featuring multiple edge aggregators. A global aggregator optimizes the collaborative utilization of anomalous knowledge from industries across various regions, enhancing anomaly detection within each industry contextually and efficiently.
- In the hierarchical FL framework, the proposed surveillance anomaly detection boosts the performance of YOLOv8n-based anomalous object detection by integrating the Flownet model for motion-based image frame extraction and pretraining the YOLOv8n model with anomalous human behaviors, thereby improving the accuracy of abnormal event detection compared to only processing entire image frames using YOLOv8n.

1.2 Paper organization

The remaining part of the paper is organized as follows. Section 2 comprehensively reviews the works related to FL-based objection detection models. Section 3 provides the system model with an architecture. Section 4 provides an overview of the proposed model and explains the mechanisms of the proposed model in detail. Section 5 shows the experimental evaluation, and section 6 concludes this paper.

2. Literature review

The related study on object detection-based surveillance using FL is discussed as follows:

Dai et al. [11] introduced an enhanced object detection algorithm focused on video key-frames to reduce latency in edge Internet of Vehicles (IoV) systems. The methodology includes a crucial coefficient and a frame similarity comparison algorithm to filter redundant frames and identify key frames for object detection. Additionally, an improved Haar-like feature-based classification algorithm is employed in the edge computation model to enhance overall detection efficiency. Liu et al. [12] presented an enhanced detection algorithm tailored for small objects using YOLOv5. This model reduced computing resources by judiciously clipping the feature map output from the large object detection layer, resulting in a more lightweight model. The study introduced an improved feature fusion method, PB-FPN, for small object detection, drawing inspiration from PANet and BiFPN. This innovation effectively enhances the algorithm's ability to detect small objects. Malburg et al. [13] introduced a framework for video-based monitoring of manufacturing processes using a physical smart factory simulation model. The study rigorously evaluates three object detection systems' efficacy in detecting workpieces and recognizing failures within the simulation model, presenting potential adaptations. These systems demonstrate reliability and provide valuable information for integration with other sensors in the IoT-based production process monitoring domain.

Cob-Parro et al. [14] designed a smart video surveillance system that employs low-power embedded devices and executes deep learning algorithms. The computer vision algorithm, optimized for surveillance, detects, counts, and tracks people's movements using MobileNet-SSD architecture. Additionally, a robust Kalman filter bank facilitates precise tracking and people counting. Based on the UpSquared2 device, the chosen edge node incorporates a vision processor unit (VPU) for accelerated AI CNN inference. Nawaratne et al. [15] presented ISTL, an Incremental Spatio-Temporal Learner, for real-time video surveillance anomaly detection and localization. ISTL employs unsupervised deep learning with active learning and fuzzy aggregation to adaptively update and distinguish evolving anomalies from normal patterns over time. Zhao et al. [16] presented an innovative Lightweight Deep Learning (DL) method termed Intelligent Edge Surveillance (LEDS), designed for applications in the Intelligent Internet of Things (IIoT). By adding depth-wise separable convolutional layers, the work in [16] adopts a multifaceted strategy to produce a lightweight neural network that mitigates the processing expenses and integrates edge and cloud computing to reduce network traffic efficiently.

Li et al. [17] developed DeepFed, a federated deep-learning method for cyber threat detection in industrial CPSs. DeepFed approach involves the designing of a new intrusion detection model using convolutional neural networks and gated recurrent units and developing an FL framework for multiple CPSs to construct intrusion detection models while preserving privacy collaboratively. Qu et al. [18] proposed a decentralized Cognitive Computing (D2C) paradigm by integrating FL and blockchain. The integration enables quick convergence with advanced verifications and member selections. Additionally, applying a Markov decision process optimizes the D2C model for accuracy and security. Huong et al. [19] presented an approach for detecting cyberattacks in Industrial Control Systems using FL-based Anomaly Detection. The architecture identifies anomalies in time series data within an IIoT-based Supervisory Management system. However, this efficiency trade-off involves increased computing resource consumption at edge devices for implementing the detection task. Liang and Wu [20] developed Edge YOLO, an Object Detection (OD) system using edge-cloud collaboration and reconstructive CNN. It solves issues of computing power reliance and uneven cloud resource distribution with a lightweight framework. Edge YOLO combines pruned feature extraction and compressed feature fusion for enhanced multi-scale prediction. Nikouei et al. [21] developed LCNN, a human object detection algorithm for edge devices, utilizing a lightweight Depth-wise Separable Convolutional network. LCNN efficiently detects pedestrians in surveillance video frames with a manageable computation workload. Implemented on a Raspberry Pi 3 with OpenCV libraries, it demonstrates satisfactory real-world performance. Chen et al. [22] introduced FedAGRU, an intrusion detection algorithm for wireless edge networks based on FL. FedAGRU updates universal learning models instead of sharing raw data and employs an attention mechanism to prioritize crucial devices, reducing communication overhead and ensuring learning convergence without unnecessary server uploads. Table 2 compares the potential factors that focus on the design of state-of-the-art anomaly detection systems.

From the review of state-of-the-art anomaly detection, it is recognized that there are still research gaps in the field of video surveillance detection of anomalies when it pertains to dealing with dynamic conditions, generalization, and computation

Table 1. Comparison of Factors involved in the Anomaly Detection Systems

Object Detection Works	Factors						
	Data Scarcity	**Anomaly Detection**	**Hybrid Model**	**Communication Cost**	**Motion Feature Analysis**	**Collaborative Training**	**Lightweight**
[11]	×	×	×	✓	×	×	×
[12]	×	×	×	×	×	×	✓
[13]	×	×	×	✓	×	×	✓
[14]	×	×	×	✓	×	×	×
[15]	×	✓	✓	×	×	×	×
[16]	×	×	×	✓	×	×	✓
[17]	×	✓	✓	✓	×	✓	×
[18]	×	✓	✓	×	×	✓	×
[19]	×	✓	✓	✓	×	✓	✓
[20]	×	×	×	✓	×	×	✓
[21]	×	×	×	✓	×	×	✓
[22]	×	✓	×	✓	×	✓	×
Proposed	✓	✓	✓	✓	✓	✓	✓

complexity. In particular, several anomaly detection researches [15, 17, 19] focused on addressing the real-time challenges; however, developing reliable video surveillance anomaly detection algorithms that dynamically address the real-time constraints without compromising the detection accuracy and speed is necessary to bridge these gaps.

Dynamicity: Traditional anomaly detection and object detection algorithms cannot respond to these changes in a diversified real-time environment, diminishing the reliability and precision performance of anomaly detection. Anomaly detection systems without motion feature analysis fail to observe temporal dynamics and contextual data about objects' activities over time. Due to the complicated or volatile nature of anomalies, a lack of contextual understanding leverages the increased false positives and false negatives. To handle this constraint, this work extracts the dynamic insights across the timeframes with motion feature analysis and collaborative knowledge from the different dynamic circumstances in distributed industries.

Generalization: Although an anomaly detection model is trained using a single surveillance dataset, it is unable to produce generalized anomalous patterns in diversified industrial circumstances. Retraining the detection model with different datasets impacts the speed of anomaly detection and increases the computational complexity in the process of detecting abnormalities in different industries. Hence, this work attempts to apply the hierarchical FL for improved generalization across various abnormal activities in distributed industries.

Computation Complexity: In large-scale surveillance systems, implementing anomaly detection requires abundant computational resources to process high-resolution video streams, which becomes challenging without affecting the speed of anomaly detection. To cope with this, the proposed anomaly detection deploys the local anomaly detection models in the edge environment and the multiple edge aggregators in the hierarchical federated settings.

3. Problem formulation and system model

This research aims to resolve the above three main research gaps - dynamicity, generalization, and computational complexity. An anomaly detection architecture is developed to achieve this objective, combining edge computing, FL, and object detection techniques specifically tailored for industrial manufacturing. The conventional automated industrial control system relies on human factors, which increases the risk of human error or deliberate abnormal actions by individuals, consequently impacting the manufacturing process. In addition, the lack of analyzing the motion features in the video sequences misleads the human anomaly detection when only relying on the object detection models. Thus, the examination of motion-based image frames in a contextual manner is imperative in addition to the extraction of knowledge regarding pixel distributions in the input video sequences.

$$F(M) = \underset{f \in Im(V)}{\operatorname{argmax}}(v_O^t) \tag{1}$$

As formulated in (1), extracting the function for motion-based image frames (f) from the images of surveillance video (V) over time (t). The maximum velocity (v) of the objects (O) in the consecutive image frames aids in recognizing the dynamics of human behaviors.

Anomaly detection examines the surveillance scene and entities or components involved, particularly in the Industry 4.0 manufacturing field. In the realm of industry, it is common to have a multitude of devices, sensors, or machines distributed across diverse locations or organizational units, leading to the increased computational strain on the anomaly detection mechanism. Hence, in order to handle this constraint, the distributed industries intricately connect with the edge server in the FL model. In this scenario, video surveillance data is collected from various geolocations to monitor and identify any abnormalities or irregularities effectively. In industrial settings, the distribution of pixel values in each image varies with the environmental factors, such as the lighting conditions, whereas the physical feature space of the indoor industrial environment becomes similar in the same kind of industries located in different regions. Hence, this work designs anomaly detection in the horizontal FL setup. Implementing horizontal FL architecture with the components of clients as the industries, edge server as the anomaly detector, and aggregator as the shared or collaborative knowledge provider. Rather than being detected from the cloud server, the anomaly detection module in the proposed system is implemented on the edge layer. The anomaly detection module in the proposed system is deployed at the edge layer instead of detected on the cloud server. It is because edge computing renders the feasibility for applications that require immediate action, such as, industrial automation and healthcare monitoring applications, leveraging the response to anomalies promptly. Moreover, FL enables the model training from the local data dispersed among various industry regions. FL ensures resilient and generalized anomaly detection by applying collaborative insights from multiple patterns. The federated setup involves the utilization of a cloud-based global model as the central monitoring entity for the locally sourced data collected from diverse localities. The centralized aggregator facilitates the exchange of parameters learned among various local industries through the local model, enabling precise anomaly detection at the edge layer rather than relying on training a global model with the sensitive data acquired from each industry. However, the increased number of communication rounds in the federated settings creates the computation and communication complexity to the anomaly detection model from the large-scale video surveillance from distributed industries.

$$O(AD)_I = \min_{i \in I} CR_I^{E,G} \tag{2}$$

$$W = arg\min_{i \in I} L\left(\overline{l_k^E}\right)_I^G \text{ ,} where\ k \subseteq I \tag{3}$$

The formulation in (2) and (3) indicates the objectives of minimizing the communication rounds ($CR_I^{E,G}$) and global loss (L_I^G) respectively during the task of Anomaly Detection (AD) in each industry. The global weights (W) are based on the average loss of edge servers ($\overline{l_k^E}$) across the industries (i), in which ‘I’ refers to the total number of distributed industries that establish communication through hierarchical FL. k, E and G denote a subset of industries connected in the edge server, edge server, and global server, respectively.

The primary components used in the design of the proposed anomaly detection are presented as follows.

Hierarchical FL: Industrial anomaly detection is a highly time-sensitive application; hence, efficiently detecting anomalous behaviors is crucial without service delay. Consequently, prioritizing the optimization of resource utilization with minimal communication overhead is highly significant in resource-constrained and heterogeneous IoT environments. To accomplish this, the hierarchical FL architecture is adopted by the proposed system, collaboratively aggregating the local model updates at various hierarchy levels. By enabling the multi-level federated aggregation of model updates in hierarchical FL [23], the proposed anomaly detection adaptively supports the heterogeneous industrial environment. Therefore, hierarchical FL is able to scale efficiently in order to accommodate distributed settings, enabling collaborative model training without overwhelming central resources. In particular, the ability to scale is the primary importance for industries engaged in extensive data collection and analysis, such as smart manufacturing, supply chain management, or energy optimization. Thus, the proposed system designs an FL-based anomaly detection system with a hierarchical FL architecture comprising the edge environment and multiple global aggregation modules. The execution of the edge server and cloud server for the aggregation depends on the computational complexity at the edge level. Consequently, the hierarchical FL offers a significant security solution for collaborative training in heterogeneous, distributed, and resource-constrained industrial settings.

Table 2. Comparison of YOLO Versions

YOLO variants	Size (pixels)	Number of Parameters (M)	FPS	FLOPs
YOLOv1	418	-	45	-
YOLOv2	416	-	40	62.94
YOLOv3	608	61.2	35	65.4
YOLOv4	608	27.6	72	59.7
YOLOv5n	640	1.9	45	4.5
YOLOv6n	640	4.3	785	11.1
YOLOv7	640	36.9	161	103.4
YOLOv8n	640	3.2	80.4	8.7

YOLO: The FL-based anomaly detection system employs an object detection model, like the YOLO framework trained with a customized surveillance anomaly dataset, to identify anomalies in indoor firms. The main objectives behind the selection of YOLO for anomaly detection in the indoor scenario are discussed as follows. i) By extracting the comprehensive features, YOLO is capable of recognizing several anomalous behaviors in a single image or frame. As a result, YOLO variants can facilitate anomaly detection in a complex and crowded indoor environment. ii) Furthermore, prompt response is a crucial component of anomaly detection, which is addressed by YOLO-based quick anomaly detection in real-time video or image frames. iii) To identify the anomalous actions carried out by individuals within the industry, the YOLO model is pretrained with data objects that highlight authorized or lawful behaviors related to the functioning of industrial components, in addition to a knowledge of abnormal human behavior.

Table 1 compares several versions of the YOLO models while testing on the PASCAL VOC dataset [24, 25]. This work selects the YOLOv8n model for object detection in the anomaly detection system from the analysis of i) computation cost based on the model parameters and Floating Point Operations (FLOPs) and ii) speed based on Frames Per Second (FPS). With its potential benefit of fast inference speeds, YOLOv8n has particular advantages for time-sensitive applications as it is optimized for real-time performance. It is capable of handling video streams quickly, allowing for the early detection of anomalies in various industries. YOLOv8n can be trained on diverse datasets to adapt to different environments and anomaly types, which enforces the precise detection of anomalies in video sequences with improved situational awareness and timely response in diverse applications.

FlowNet: The main application of FlowNet is optical flow estimation, which estimates the motion vectors between successive frames in a video stream. FlowNet is able to forecast the intense optical flow field that occurs between successive frame pairs when the flow field mentions the motion of the pixels in the sequential image frames. FlowNet's optical flow information has been an important source of input for conventional object identification models like CNN and region-based detectors, particularly YOLO, providing additional benefits. In video surveillance, optical flow data offers insightful context on motion dynamics that enhances the resilience and accuracy of object detection systems when there are dynamics in the objects.

The proposed architecture comprises three main layers: data, edge, and cloud, as illustrated in Fig. 1.

Data Layer: In a three-tier architecture, the lower layer refers to the data layer that consists of several industrial devices, such as sensors, actuators, and computing units, placed in various sections of the industrial facility. The data layer is responsible for monitoring and recording the videos using the IoT devices, and the object detection model mounted on the edges gained data-level knowledge from the monitored data.

Edge Layer: The main objectives of the edge layer design in the proposed anomaly detection system target to optimize the real-time quick response, scalability, and communication overhead across distributed industrial environments. In particular, the edge layer is responsible for executing surveillance anomaly detection mechanisms with the integration of different local models for dispersed industries. The anomaly detection mechanism is a design of a local YOLO v8-assisted object detection model based on its locally collected industrial information. In the hierarchical FL-enabled proposed system, edge aggregation is also performed with the concept of FedAvg for the different local models at the lower layer of the edge in which the edge aggregator is located on the upper layer of the edge. Moreover, the cloud server updates the local model with global parameters based on prevalent knowledge of anomalous behavioral objects across multiple industries. The hierarchical aggregation process iteratively continues over the hierarchical levels until convergence is attained based on the predefined stopping criterion.

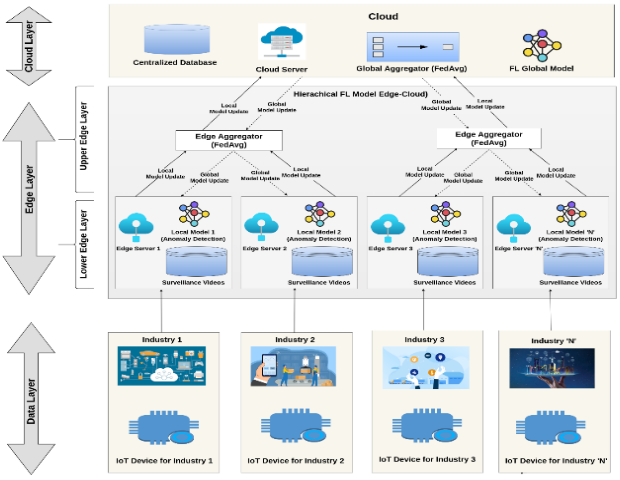


Figure. 1 The architecture of FL-based Anomaly Detection in Industry 4.0

Cloud Layer: It comprises a server, database, and FL aggregator. The cloud server generates a comprehensive FL global model by aggregating the local model parameters obtained from different edges through FedAvg. The global model performance is cooperatively improved by the hierarchical level of multiple intermediate nodes and central servers that aggregate the updates received from various industries. To produce greater accuracy in the global model, a greater number of interactions between the cloud server and edges are essential. In order to enable anomaly detection in every industry, even in circumstances with insufficient abnormal object patterns, FL conducts a distributed model training procedure in which numerous edge devices collaboratively improve a shared global model without raw data exchange.

4. Proposed system

The primary objective of the proposed work is to determine the abnormal events conducted by humans in Industry 4.0 by introducing a novel FL-enabled object detection-based surveillance strategy in the edge layer. The proposed anomaly detection system incorporates three main phases: data collection and feature extraction, edge-assisted FL, and YOLO v8-

based object detection. Initially, the most pertinent features are extracted as image frames by preprocessing the input dataset. Secondly, at the edge, the FL-based surveillance analysis model enhances object detection learning strategies by generating extensive global knowledge while maintaining privacy. The proposed method applies the hierarchical FL that enables the multi-level aggregation at the edge and cloud server rather than transferring the local model parameters from the edge to the remote cloud for the federated aggregation. Finally, by combining the strengths of Flownet and YOLO v8 models for object identification in the hierarchical FL model, the proposed anomaly detection strategy identifies anomalies related to human behavior in situations. In particular, using pretrained knowledge from the YOLOv8n model customized by human anomalous behaviors enforces the anomaly detection model to trigger an emergency alarm. As a result, the design of hierarchical FL and edge intelligence leverages the execution of the proposed system under minimal service delay without impacting the quality of anomaly detection.

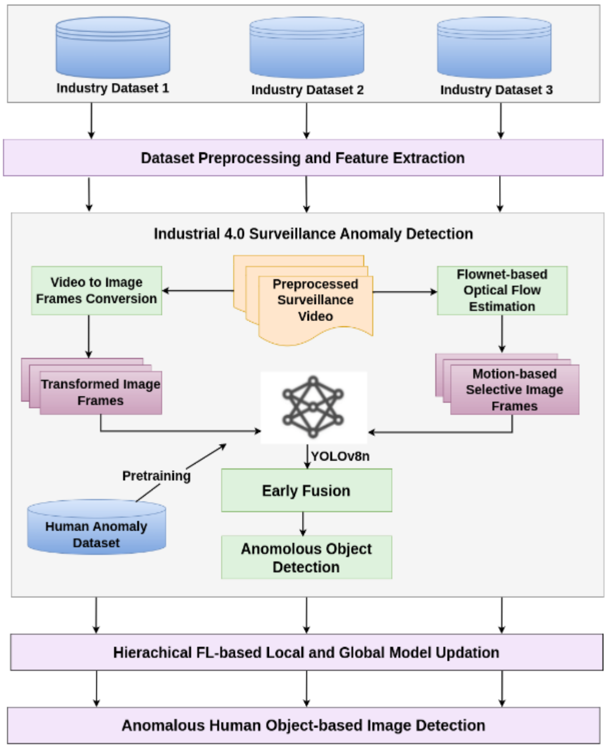

Industry Dataset 1
Industry Dataset 2
Industry Dataset 3
Dataset Preprocessing and Feature Extraction
Industrial 4.0 Surveillance Anomaly Detection
Video to Image Frames Conversion
Preprocessed Surveillance Video
Flownet-based Optical Flow Estimation
Transformed Image Frames
Motion-based Selective Image Frames
YOLOv8n
Pretraining
Early Fusion
Human Anomaly Dataset
Anomolous Object Detection
Hierachical FL-based Local and Global Model Updation
Anomalous Human Object-based Image Detection

Figure. 2 Proposed Anomaly Detection Methodology

4.1 Surveillance data feature extraction

To implement the FL model for the distributed industries, the proposed system needs to be evaluated on the datasets that are timestamp-tagged surveillance videos, including normal and abnormal human activities gathered from the indoor industrial scenario. To detect the anomaly in the manufacturing industry, the proposed system examines the interactions between human and machine or industrial components, facilitating the spotting of anomalous human behavior in the object detection outcome. For the purpose of effectively analyzing and interpreting the data gathered from surveillance systems, surveillance data feature extraction is essential.

Owing to the lack of industrial surveillance videos for anomaly detection, this work utilizes several datasets, such as the Avenue dataset [26], UCF-Crime dataset [27], UMN-Crowd11 dataset [28], and SVIP dataset [29] with multiple surveillance videos to demonstrate the anomalous object detection in the context of the human-involved manufacturing industry. Even though the aforementioned surveillance datasets are not particularly developed from indoor industrial spaces, the image frames and human activities in these datasets leverage anomalous human object detection. The Avenue dataset [26] comprises videos captured from indoor spaces and streets, widely exploited for anomaly detection. UCF-Crime dataset [27] is also used for the anomaly detection task consisting of various human-involved crime scenarios of video clips, such as the assault, robbery, and burglary recorded from the surveillance cameras. UMN Crowd11 dataset [28] contains the motion patterns of the crowd with the composition of 6000 video sequences. Shanghaitech Vision and Intelligent Perception (SVIP) dataset [29] comprises the human actions in surveillance footage with 130 abnormal events over 270,000 training frames, widely utilized for anomaly detection tasks. In conclusion, this work employs the SVIP dataset as the human anomalous dataset for pretraining the YOLOv8n model. The remaining three datasets [26-28] are considered local for three industrial environments.

During the video preprocessing, the proposed approach enhances the visual tracking of physical parts by implementing depth-based segmentation and background removal. The process of extracting relevant information from acquired video frames for anomaly detection is a major component of feature extraction, which is used extensively in video surveillance. In order to achieve promising outcomes for the stacked multiple image frames over time, the proposed approach applies a CNN-based model to spatially extract object patterns from the surveillance video due to the effective feature extraction providing situational awareness to the anomaly detection system. Motion pattern analysis, in particular, is a component of the feature extraction method that helps to identify anomalous human actions in video frames accurately. In the video sequences, motion features involve the optical flow, acceleration, speed, and direction, which assist in characterizing the dynamics of the objects. Therefore, the proposed approach aims to automatically analyze the optical flow in the visual frames for motion tracking, thereby rendering it simple to identify anomalous objects in the surveillance footage. Instead of providing all the image frames to the YOLO model, the proposed approach extracts the motion features using the Flownet model briefly discussed by Savian et al. [30]. Subsequently, the motion-based image frames are provided to the YOLO model and the raw inputs. In the proposed system, the Flownet model is integrated with the YOLO model as the local anomaly detection model in the FL environment.

4.2 Edge-assisted federated learning

To address data scarcity and handle heterogeneity in the manufacturing industry, the proposed approach applies the edge-assisted hierarchical FL model that significantly mitigates the communication cost with the globally shared intelligence. The edge-assisted FL facilitates cooperative model training over-dispersed edge devices while protecting data privacy and reducing communication overhead by integrating the FL paradigm with edge computing infrastructure. The proposed surveillance anomaly detection is implemented on the edge network in the FL infrastructure with the components of a local model, edge aggregator, and cloud or global aggregator. The proposed edge-assisted FL model addresses the privacy and cost constraints while storing the surveillance videos of each industry in the centralized cloud storage. To avoid transferring videos for the global model training it is achieved by building an object detection model locally on the associated edge server. By utilizing the potential benefits of edge computing, the proposed approach implements hierarchical FL with minimal delay and low cost by using partitioned edge networks as the upper and lower layers. In particular, time-critical applications, such as surveillance anomaly detection, are increasingly beneficial by the hierarchical modeling of FL. The proposed FL architecture aims to resolve the shortcomings in a single aggregator for the trained local models by collectively integrating the multiple local models for the edge-level aggregator and the cloud-level aggregator. The proposed federated architecture is hierarchical, with several layers of edge devices to train anomaly detection models cooperatively while maintaining data privacy. Lower-layer edge devices deploy a lightweight learning model to interpret locally accessible surveillance video, extract pertinent features, and identify anomalies. The global anomaly detection model is then improved by combining insights from a two-layer edge network in which the local model parameters are subsequently aggregated at the upper edge layer, improving anomaly detection accuracy. In the edge network, edge aggregator placement is based on the number of requests and characteristics of the communication network in a particular region. As illustrated in Fig. 1, in the hierarchical FL model, the proposed anomaly detection module is located on the lower or bottom layer of the edge network associated with the federated edge aggregator located on the higher or top layer of the edge network.

Moreover, the global federated aggregator is located on the cloud, interacts with the hierarchical aggregation in the edge layer, and ensures the time-efficient anomaly detection mechanism. The proposed local model, an object detection-based anomaly detection model, initially learns the local industry dataset in the corresponding edge server. Subsequently, the edge aggregator is responsible for amalgamating the insights obtained from the local models of several businesses by exchanging local model parameters rather than disclosing the raw surveillance video input. In the edge layer, the proposed approach initiates a relevant local model update based on the model's parameters in the edge aggregator. Moreover, the multiple local models trained on data from various industries by the numerous edge aggregators in a hierarchical FL structure enforce the global aggregation on the cloud server. During the aggregation, the FedAvg method collaboratively learns the model parameters and weight to update the edge aggregator in the top layer of the edge network and the global aggregator in the cloud network. Finally, the proposed approach reduces service latency. It enhances anomaly detection quality for time-sensitive anomaly detection in the manufacturing sector by updating local models across industries using hierarchically shared global parameters.

4.3 YOLO-based anomalous object detection

The primary goal of this work is to develop an effective anomaly detection system that efficiently extracts temporal and spatial information from surveillance videos by integrating the FlowNet and YOLOv8 models. In the unified anomaly detection system, the YOLOv8 component offers object-based features that capture spatial relationships and object properties, and the FlowNet component offers motion-based features that observe temporal dynamics and motion patterns. The proposed anomaly detection integrates Flownet and YOLOv8 model for object detection in the hierarchical FL architecture, implemented in the edge layer. Anomaly detection effectively finds and categorizes abnormal activities or patterns in real-time surveillance video streams by utilizing the distinctive objects extracted from the You Only Look Once (YOLO) object identification framework. Owing to resource limitations in the edge environment, the proposed approach targets to create a lightweight object detection model with the minimal possible YOLOv8n model parameters comprising 3.2M parameters and recognizing the 640×640 pixels of image frames. In the series of object detectors, YOLOv8n is the most recent variant with the novelty of enhancement and features in its architecture. Compared to other YOLO models, YOLOv8 detects objects even in real-time crowd density with the proficiency of recognizing abnormal activities.

To customize the YOLOv8n model for surveillance anomaly detection, the proposed approach trains the object detection model with the SVIP dataset [29] that contains the human abnormal behaviors, facilitating anomalous human activities through YOLOv8n-based object detection. The depth-wise separable convolution layers-enabled YOLO model leverages fast anomalous object detection to ensure lightweight real-time object detection. The enhanced anomalous object detection described in Fig. 3 is enriched by the Flownet model [30], which assists in minimizing or discarding the repetitive video frames in the surveillance video activities and provides the filtered input into the YOLOv8n model. Consequently, the proposed object detection model accurately recognizes the abnormal activities of humans from the filtered motion-based video frames alone and reduces the computation complexity while understanding human activities. Motion characteristics capture an object's dynamic activity all over time, thereby providing temporal context. Systems for detecting anomalies can distinguish between standard and anomalous object activities or interactions in the environment according to this temporal information. Moreover, the surveillance video frames are provided as the input to the YOLOv8n model in parallel and thus, both the Flownet-assisted YOLOv8n features and raw input-based YOLOv8n features are early fused before the object detection in its architecture.

In the proposed anomaly detection, the design of the lightweight object detection model is enhanced with the MobileNetv3 model for backbone feature extraction to further increase the speed of the computations due to the deployment on the resource-constrained edge network. MobileNetV3 is a promising solution that offers deployment scalability in distributed industrial environments due to its flexibility and adaptability to different datasets and object detection tasks. Combined with its lightweight architecture, MobileNetV3 performs better comparable to extremely complicated models in object detection tasks and reaches competitive accuracy levels. In MobileNetV3, the significance of model parameters and advanced features, such as the squeeze-and-excitation blocks and inverted residuals, aids in reducing the computation cost while increasing accuracy. By modeling the h-swish activation function, the MobileNetV3-integrated YOLOv8n model reduces the computations and improves the model performance. Therefore, the proposed object identification precisely identifies anomalies pertaining to humans without Introducing the computing burden. By applying the edge-assisted network, the federated architecture and YOLOv8n model jointly provide efficient and real-time anomaly detection in the industrial sector. From the analysis of anomalous image frames detected by the object detection model, higher anomaly scores indicate a higher probability of abnormal behavior, accomplished by the thresholding method that decides whether the detected object is anomalous or normal. In surveillance anomaly detection, the thresholding method is based on the examination of anomaly scores over the time frame. Finally, the anomaly detection model generates alerts or triggers the automatic response systems in the industry. Algorithm 1 describes the overall steps involved in the proposed surveillance anomaly detection.

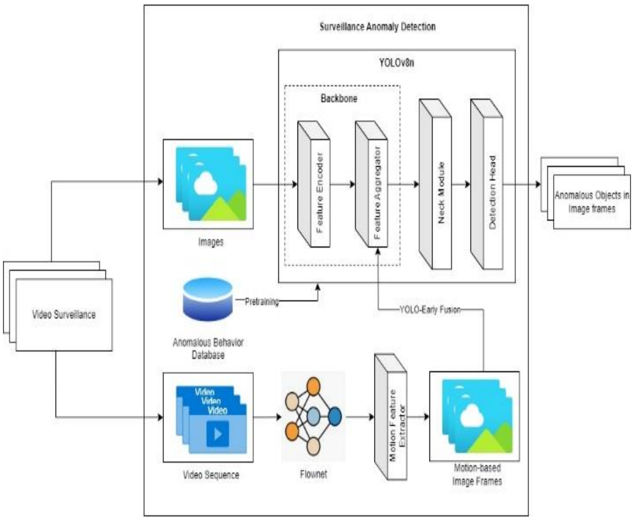
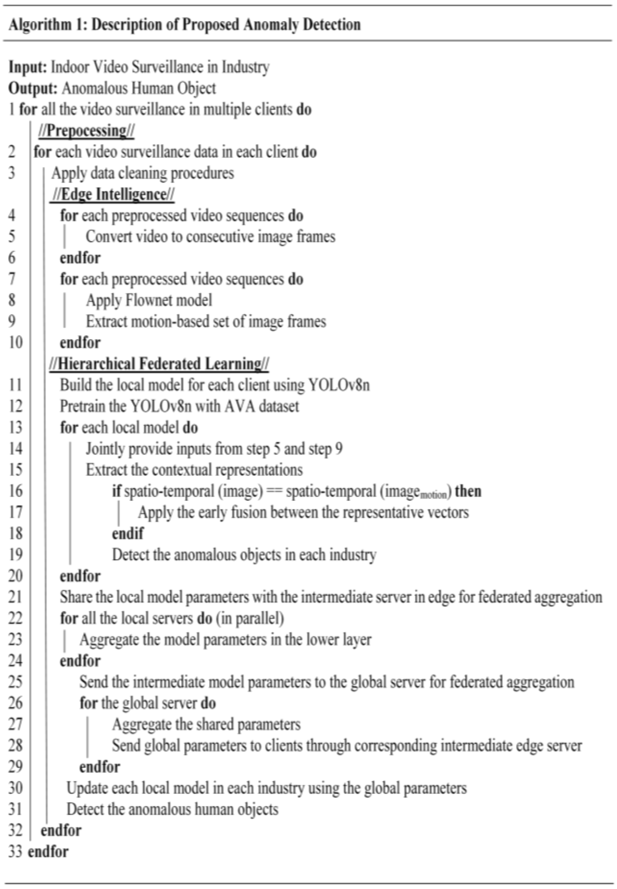


Figure. 3 Local Model in Federated-Enabled Proposed Anomaly Detection



```
Algorithm 1: Description of Proposed Anomaly Detection

Input: Indoor Video Surveillance in Industry
Output: Anomalous Human Object
1  for all the video surveillance in multiple clients do
       //Prepocessing//
2      for each video surveillance data in each client do
3          Apply data cleaning procedures
           //Edge Intelligence//
4          for each preprocessed video sequences do
5              Convert video to consecutive image frames
6          endfor
7          for each preprocessed video sequences do
8              Apply Flownet model
9              Extract motion-based set of image frames
10         endfor
           //Hierarchical Federated Learning//
11         Build the local model for each client using YOLOv8n
12         Pretrain the YOLOv8n with AVA dataset
13         for each local model do
14             Jointly provide inputs from step 5 and step 9
15             Extract the contextual representations
16                 if spatio-temporal (image) == spatio-temporal ($image_{motion}$) then
17                     Apply the early fusion between the representative vectors
18                 endif
19                 Detect the anomalous objects in each industry
20         endfor
21         Share the local model parameters with the intermediate server in edge for federated aggregation
22         for all the local servers do (in parallel)
23             Aggregate the model parameters in the lower layer
24         endfor
25             Send the intermediate model parameters to the global server for federated aggregation
26             for the global server do
27                 Aggregate the shared parameters
28                 Send global parameters to clients through corresponding intermediate edge server
29             endfor
30         Update each local model in each industry using the global parameters
31         Detect the anomalous human objects
32     endfor
33 endfor
```

5. Experimental evaluation

To evaluate surveillance anomaly detection, this work experiments with the object detection model in the FL setup for the proposed algorithm by comparing several existing object detection, E1 and E2 [16, 12], FL works, E3 and E4 [17, 18], and surveillance anomaly detection research works, E5 and E6 [19, 15] respectively. In the evaluation scenario, the comparative existing works are referred to as E1, E2, E3, E4, E5, and E6.

5.1 Implementation setup

The experimental model is implemented using Python language with the version of Python 3.8 in the Ubuntu 18.04 Operating system for conducting the YOLOv8 model and FL experiments. To train the YOLOv8n model with surveillance anomalies, the experimental model exploits the video surveillance datasets [26-28] as indoor human anomaly videos for three industries in the federated settings. The hierarchical federated setting is implemented with 5 communication rounds for 3 clients. To assess the surveillance anomaly detection, the experimental model employs precision, recall, mean average precision (mAP), and accuracy metrics to validate the object and anomaly detection in the surveillance videos. Precision is the percentage of detected objects or anomalies, the recall is the percentage of accurately identified anomalies based on the identified bounding boxes in the ground truth.

Table 3. Implementation Parameters of Object and Anomaly Detection Models.

Parameter	**Values**			
	Proposed	**Liu et al. (2022) [12]**	**Nawaratne et al., (2020) [15]**	**Zhao et al. (2020) [16]**
Model	YOLOv8	YOLOv5	Spatio Temporal Autoencoder	Tiny-YOLO
Image Size	640×640	640×640	224×224	256×256
Optimizer	Adam	Adam	Adam	Adam
Activation Function	Relu, Sigmoid	Relu, Sigmoid	Relu, Sigmoid	Relu, Sigmoid
Loss Function	Binary Cross Entropy	Binary Cross Entropy	Binary Cross Entropy	Binary Cross Entropy
Learning Rate	0.001	0.001	0.001	0.001
Epochs	10	10	10	10
Batch Size	32	32	32	32
Backbone	MobileNetV3	Spatial pyramid pooling at the end of the backbone	-	MobileNetV2-SSD

Moreover, to precisely assess the ability of anomalous activity detection by the proposed model, the experimental model computes the anomaly score for each detected image frame using the following Eqs. (4) and (5).

$$PSNR\big(A(I), P(I)\big) = 10\log_{10}\frac{\left[Max.score_{P(I)}\right]^2}{\frac{1}{N}\sum_{i=1}^{N}\left(score(i)_{A(I)} - score(i)_{P(I)}\right)^2} \tag{4}$$

$$AS(t) = \frac{PSNR_t - min(PSNR)}{max(PSNR) - min(PSNR)} \tag{5}$$

Table 4. Implementation Parameters of Federated Settings.

Parameter	Values			
	Proposed	Li et al., (2021) [17]	Qu et al., (2021) [18]	Huong et al., (2021) [19]
Federated Architecture	Hierarchical FL	As proposed in [17]	As proposed in [18]	As proposed in [19]
Model Architecture	Yolov8	Yolov8	Yolov8	Yolov8
Communication Rounds	5	5	5	5
Number of Clients	3	3	3	3
FedAvg	Yes	Yes	Yes	Yes
Epochs	10	10	10	10
Batch Size	32	32	32	32

To compute the Anomaly Score (AS) for each image frame (i), the experimental model measures the Peak Signal-to-Noise Ratio (PSNR) between the actual (A) and predicted (P) image frames based on the score returned by the anomaly detection model. In (4), ‘i’ denotes each image frame in the total ‘N’ number of image frames. To normalize the PSNR of the image frames in a video, Eq. (5) computes the normalized anomaly score. Accordingly, the experimental model exemplifies the performance of the proposed algorithm in terms of average anomaly score per video segment. Table 1 depicts the training parameters used for the proposed and existing object and anomaly detection model for the implementation of test datasets.

Furthermore, to implement the federated model in the proposed system, the experimental framework provides the federated parameter settings of the proposed approach and several existing federated approaches in Table 2. The comparative both the existing federated and proposed approaches are evaluated on those above all the three datasets.

As mentioned in Table 2, the experimental model conducts the experiments for the existing FL architectures [17-19] with the same evaluation scenarios. The experimental model evaluates the effectiveness of the federated architectures and workflow in the corresponding works [17-19] with the deep learning architectures of Yolov8 as adopted in the proposed system. Even though the existing federated approaches [17-19] employed different benchmark datasets, this experimental model utilizes the three benchmark above surveillance video datasets for the input of YOLOv8 models in their federated architectures, illustrating the ability of anomaly detection by each federated approach. For the evaluation of work in [17], the workflow of the DeeFed scheme is considered as the existing algorithm with the YOLOv8 deep learning architecture rather than the CNN-GRU-based intrusion detection model. In contrast to the privacy assessment, the experimental model evaluates the work in [18] in terms of the FL-based cognitive computing algorithm implementation. During the evaluation of work in [19], the experimental model implements the federated edge architecture with the Yolov8 model as the local model instead of the VAE-LSTM model.

5.2 Results and discussion

From the analysis of Fig. 4, the proposed object detection model obtains higher precision, recall, and mAP while testing on the surveillance videos of multiple industries.

5.2.1. Evaluation of object detection models

Compared to the various YOLO models, such as the YOLOv4 and YOLOv5, the edge YOLOv8 model outperforms the object detection performance in the video frames. All the comparative models and research works [12, 16] are evaluated under the same experimental settings, including the optimizer, activation function, loss function, epochs, batch size, and learning rate. The comparative research only varied in their corresponding algorithms and input image size, as mentioned in Table 1.

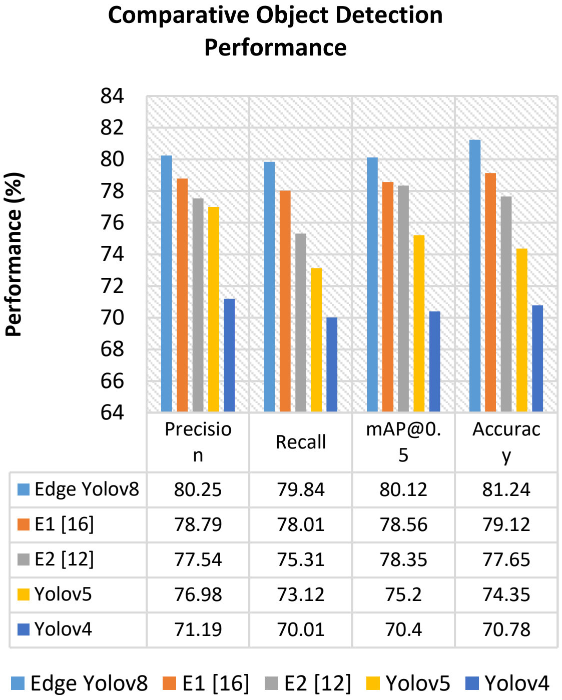
As shown in Fig. 4, the edge yolov8 model precisely recognizes the objects for the test dataset with the potential advantage of enhanced feature extraction and computation capabilities in the object detection architecture. The results depicted in Fig. 4 are obtained from the implementation of all the comparative algorithms on the Avenue dataset [26]. Conventional object detection research works [12, 16] have yielded results on various video datasets. To compare the object detection performance, the experimental model implements the object detection algorithms [12, 16] for the Avenue dataset and measures the performance metrics. As a result, the edge YOLOv8 model accomplishes 81.24% accuracy values, which is 3.59% higher than the comparative object detection model [12]. Even though the

Received: April 10, 2024. Revised: May 26, 2024.

comparative model [12] applies the YOLOv5 model for object detection, the edge YOLOv8 model outperforms the human recognition performance even in a crowded environment, particularly in surveillance videos. In addition, ensuring generalization ability is critical to analyzing its object detection performance on a variety of datasets and scenarios due to the lack of training in the model with global or collaborative knowledge. Also, the comparative edge intelligence in [16] applies the tiny YOLO with the integration of MobileNetV2-SSD. However, the YOLOV8 model outperforms the object detection performance when deployed on the edge environment. It is because edge intelligence significantly improves accuracy and efficiency by enabling contextual edge-cloud collaboration and dynamic adaptation to the input data, leveraging the extraction of actionable insights for accurate decision-making. Hence, this work selects the YOLOv8 model as the core model for object detection among the abnormal activities of human objects due to its superior performance in object detection from video frames and the integration of edge intelligence.

5.2.2. Impact of federated learning on anomaly detection

Table 3 presents the performance of the proposed surveillance anomaly detection with the comparison of the existing FL-based anomaly detection approach [17] and centralized approach while testing on the Avenue dataset [26], UCF-Crime dataset [27], and UMN Crowd11 dataset [28] with 3 clients. From the results mentioned in Table 3, it is determined that the proposed hierarchical FL approach outperforms the traditional federated approach [17] and the centralized approach in anomaly detection with improved accuracy and mAP@0.5. In addition, the work in [18] enables quick convergence with optimal verification in FL settings and obtains 2.37% higher accuracy than the E3 [17]. Despite this, the hierarchical FL in the proposed approach outperforms the existing federated approaches [17, 18] by contextually integrating the insights from the diversified industries in the multiple edge aggregators, resulting the 88.95% accuracy.



	Precision	Recall	mAP@0.5	Accuracy
Edge Yolov8	80.25	79.84	80.12	81.24
E1 [16]	78.79	78.01	78.56	79.12
E2 [12]	77.54	75.31	78.35	77.65
Yolov5	76.98	73.12	75.2	74.35
Yolov4	71.19	70.01	70.4	70.78

Figure. 4 Assessment of Object Detection Performance

Table 5. Comparative Performance of Anomaly Detection Models

Comparative Models	Average Anomaly Detection Performance (%)			
	Precision	Recall	Accuracy	mAP @0.5
Centralized	80.25	79.59	80.05	80.11
Li et al., (2021), E3 [17]	84.01	83.98	84.41	83.88
Qu et al., (2021), E4 [18]	86.13	85.17	86.78	86.04
Huong et al., (2021), E5 [19]	87.78	86.34	87.05	86.95
Proposed	87.76	86.34	88.95	87.19

As presented in Table 3, the experimental model partially implements the existing federated approaches in the context of anomaly detection for the surveillance videos. Although the research [17, 19] fails to target the communication between the edge and cloud as well as the edge and IoT devices during the federated execution, it leads to increased bandwidth resource consumption and affects communication efficiency. In addition, the aggregator in the federated environment focused on the same data distribution or patterns from each local model affects anomaly detection when there are inherently varied anomaly patterns in the input data. In contrast, the proposed approach applies the hierarchical FL with multiple edge aggregators and leverages the accurate understanding of the input patterns for the anomaly detection task.

5.2.3. Evaluation of the proposed anomaly detection

Table 4 compares the impact of each task and model involved in the proposed system on improving the performance of anomalous object detection while testing on three datasets above for the centralized and federated settings. In Table 4, the evaluation results of the centralized settings are obtained for the Avenue dataset, whereas the federated settings are also obtained for the Avenue dataset with the collaborative integration of UCF-crime and UMN-Crowd11 datasets during federated aggregation. Moreover, the tasks of YOLOv8n, YOLOv8n+MobileNetv3, and YOLOv8n + Flownet model belong to the centralized settings, whereas other tasks are under the federated settings. Compared to the YOLOv8n model, integrating the Flownet and Hierarchical FL (HFL) model with the surveillance anomaly detection system yields better true positive rates at 85.18% and 86.34%, respectively. Even though the YOLOv8n model accurately recognizes the objects, integrating collaborative knowledge from diverse environments and motion feature-based image frame analysis enforces anomalous object detection in video surveillance with improved accuracy. Moreover, the HFL enables the communication and cooperation between edge devices, enhancing anomalous object detection performance and generalization. The proposed approach improves the anomaly detection accuracy by jointly utilizing the diversified local model updates at several hierarchical levels in addition to the abnormal behavior pretraining in the YOLOv8n model. Also, it minimizes the communication rounds by achieving higher accuracy in minimal communication rounds rather than executing the multiple communication rounds.

5.2.4. Evaluation of anomaly detection

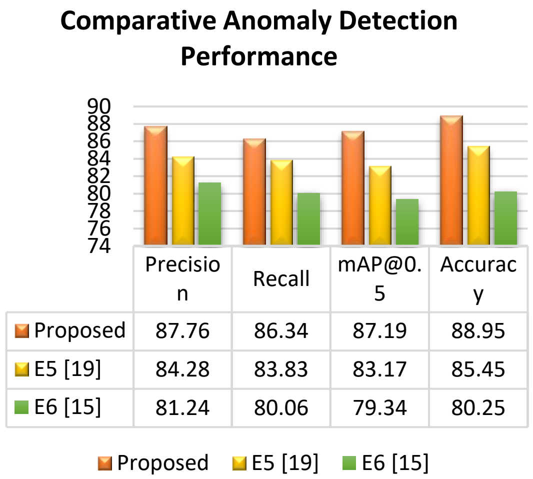
Fig. 5 compares the proposed anomaly detection performance with the existing surveillance anomaly detection approach [15, 19] in which the Avenue dataset evaluates existing [15, 26] in this experiment. In contrast, the time-series industry datasets implement the evaluation of the E5 approach [19]. To unify the comparative evaluation of the proposed approach and the existing works [15, 19], the experimental model utilizes the Avenue dataset as the test dataset. During the proposed model evaluation, the anomaly detection performance of the Avenue dataset is illustrated in Fig. 5 by collaboratively utilizing the UCF-crime and UMN-Crowd11 dataset in federated settings. The proposed approach yields an accuracy of 88.95% and mAP as 87.19%, which is 8.7% and 7.85% higher than the comparative anomaly detection model [15]. The proposed approach integrates the FL model with the YOLOv8n model for accurate object detection. Also, the Flownet-based motion-video frames greatly assist the recognition of abnormal activities by humans in the industrial environment.

Table 6. Comparative Performance of Proposed Anomaly Detection Tasks

Tasks	Average Anomaly Detection Performance (%)			
	Precision	Recall	Accuracy	mAP@0.5
Yolov8n	78.33	78.53	78.24	78.13
Yolov8n + MobileNet V3	79.37	79.75	79.58	79.01
Yolov8n + FlowNet	80.25	79.59	80.05	80.11
Yolov8n + FlowNet + FL	85.98	85.18	85.33	85.76
Proposed (Yolov8n + FlowNet + HFL)	87.76	86.34	88.95	87.19

Even though the existing 6 model in [15] analyzes the spatial and temporal features in the surveillance videos, the lack of YOLO-based object detection fails to capture the inherent pattern changes during the pixel-level analysis. Also, the comprehensive is provided to the object detection model in a particular context by training the model with the videos of human anomalies dataset. Although the anomaly detection model in [19] recognizes the anomalous activities in the smart manufacturing industry, it becomes inaccurate when adopting for the surveillance videos in the Avenue dataset, even when adopting the YOLOv8 as the model architecture in the federated model of [19] and thus, accomplishes a 3.77% higher true positive rate than the E5 [15]. The collaborative knowledge shared by the hierarchical FL model and pretraining the YOLOv8n model with abnormal activity in the proposed system enforces the accurate detection of anomalies in each industry from the video surveillance with 87.19% mean average precision value, even when there are inadequate surveillance videos with anomalous patterns.

Fig. 6 compares the proposed surveillance anomaly detection with the E6 [15] in the perspective of examining the average anomaly score of the



	Precision	Recall	mAP@0.5	Accuracy
Proposed	87.76	86.34	87.19	88.95
E5 [19]	84.28	83.83	83.17	85.45
E6 [15]	81.24	80.06	79.34	80.25

Figure. 5 Assessment of Surveillance Anomaly Detection Performance

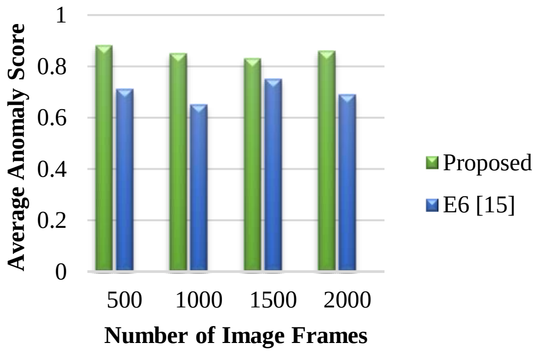


Figure. 6 Assessment of Anomaly Score

predicted anomalous image frames in the video segment. From the analysis of Fig. 6, the anomaly detection performance within the bounding boxes of the image frame rather than assessing the number of detected anomalous image frames. Moreover, to prove the scalability of the anomaly detection performance, the average anomaly score is evaluated for different numbers of image frames in each video. Consequently, the proposed approach maintains the anomaly score with an average of 0.85 even when there is a huge number of image frames in the surveillance video segment, accomplished by edge intelligence and hierarchical FL.

6. Conclusion

This work suggested a surveillance anomaly detection model for the manufacturing industry, integrating a hierarchical FL model and enhanced YOLOv8n-based object detection by the Flownet. The proposed anomaly detection model is deployed within the edge network infrastructure, adopting a federated setting for enhanced performance. This architecture entailed the placement of both the local model and multiple FL-based edge aggregators directly on the edge layer. This setup is precisely designed to facilitate accurate decision-making while minimizing communication rounds, thus optimizing the efficiency of anomaly detection processes. Moreover, by leveraging this distributed approach, the system empowered earlier detection and warning of anomalous activities within the industry, contributing to improved security and proactive risk mitigation measures. Moreover, the flow net-assisted motion-based image frames extraction enhanced the learning ability of the YOLOv8n model in the context of anomalous behavior learning over the consecutive image frames leveraged the accurate detection of anomalies in the surveillance videos. Thus, this work ensured that the proposed surveillance anomaly detection is an accurate recognition of 88.95% accuracy for Industry 4.0 applications.

Conflicts of Interest

Declare conflicts of interest or state “The authors declare no conflict of interest.” Authors must identify and declare any personal circumstances or interest that may be perceived as inappropriately influencing the representation or interpretation of reported research results.

Author Contributions

Conceptualization: Ibrahim Alnajjar and Laiali Almazaydeh came up with the main idea and framework for the study, setting the research objectives and goals. Methodology and Software: Ibrahim Alnajjar designed the research methodology, while Ali Odeh developed and implemented the software models, ensuring the technical aspects were correctly executed. Validation and Formal Analysis: Ibrahim Alnajjar, Laiali Almazaydeh, and Anas Salameh validated the results, cross-checking for accuracy, while Ibrahim Alnajjar conducted the formal analysis and interpretation of data. Writing-review and editing: The review and editing of the manuscript were done by Ibrahim Alnajjar, Khalid Alqarni and Anas Ban Atta ensuring the final document was polished and accurate. Funding Acquisition: The funding for the research was secured by Anas A. Salameh, who ensured the financial support necessary for the study.

Acknowledgments

This study is supported via funding from prince Sattam bin Abdulaziz University project number (PSAU/2023/R/1445).

References

[1] L. L. Dhirani, E. Armstrong, and T. Newe, "Industrial IoT, cyber threats, and standards landscape: Evaluation and roadmap", *Sensors (Basel)*, Vol. 21, No. 11, p. 3901, 2021.

[2] A. Jamwal, R. Agrawal, M. Sharma, and A. Giallanza, "Industry 4.0 technologies for manufacturing sustainability: A systematic review and future research directions", *Appl. Sci. (Basel)*, Vol. 11, No. 12, p. 5725, 2021.

[3] T. Qiu, J. Chi, X. Zhou, Z. Ning, M. Atiquzzaman, and D. O. Wu, "Edge computing in industrial internet of things: Architecture, advances and challenges", *IEEE Commun. Surv. Tutor.*, Vol. 22, No. 4, pp. 2462–2488, 2020.

[4] S. Hamdan, M. Ayyash, and S. Almajali, "Edge-computing architectures for Internet of Things applications: A survey", *Sensors (Basel)*, Vol. 20, No. 22, p. 6441, 2020.

[5] W. Xiang, K. Yu, F. Han, L. Fang, D. He, and Q. -L. Han, "Advanced Manufacturing in Industry 5.0: A Survey of Key Enabling Technologies and Future Trends", *IEEE Transactions on Industrial Informatics*, Vol. 20, No. 2, pp. 1055-1068, 2024, doi: 10.1109/TII.2023.3274224.

[6] S. H. Alsamhi, A. V. Shvetsov, A. Hawbani, S. V. Shvetsova, S. Kumar, and L. Zhao, "Survey on Federated Learning enabling indoor navigation for industry 4.0 in B5G", *Future Gener. Comput. Syst.*, Vol. 148, pp. 250–265, 2023.

[7] E. T. M. Beltrán et al., "Decentralized Federated Learning: Fundamentals, state of the art, frameworks, trends, and challenges", *arXiv.2211.08413*, 2022.

[8] J. Zhou et al., "A survey on federated learning and its applications for accelerating industrial Internet of things", *arXiv:2104.10501v1*, 2021.

[9] J. Wen, Z. Zhang, Y. Lan, Z. Cui, J. Cai, and W. Zhang, "A survey on federated learning: challenges and applications", *Int. J. Mach. Learn. Cybern.*, Vol. 14, No. 2, pp. 513–535, 2023.

[10] A. N. Jahromi, H. Karimipour, and A. Dehghantanha, "An ensemble deep federated learning cyber-threat hunting model for Industrial Internet of Things", *Comput. Commun.*, Vol. 198, pp. 108–116, 2023.

[11] C. Dai, X. Liu, W. Chen, and C.-F. Lai, "A low-latency object detection algorithm for the edge devices of IoV systems", *IEEE Trans. Veh. Technol.*, Vol. 69, No. 10, pp. 11169–11178, 2020.

[12] H. Liu, F. Sun, J. Gu, and L. Deng, "SF-YOLOv5: A lightweight small object detection algorithm based on improved feature fusion mode", *Sensors (Basel)*, Vol. 22, No. 15, p. 5817, 2022.

[13] L. Malburg, M.-P. Rieder, R. Seiger, P. Klein, and R. Bergmann, "Object detection for smart factory processes by machine learning", *Procedia Comput. Sci.*, Vol. 184, pp. 581–588, 2021.

[14] A. C. Cob-Parro, C. Losada-Gutiérrez, M. Marrón-Romera, A. Gardel-Vicente, and I. Bravo-Muñoz, "Smart video surveillance system based on edge computing", *Sensors (Basel)*, Vol. 21, No. 9, 2021.

[15] R. Nawaratne, D. Alahakoon, D. De Silva, and X. Yu, "Spatiotemporal anomaly detection using deep learning for real-time video surveillance", *IEEE Trans. Industr. Inform.*, Vol. 16, No. 1, pp. 393–402, 2020.

[16] Y. Zhao, Y. Yin, and G. Gui, "Lightweight deep learning based intelligent edge surveillance techniques", *IEEE Transactions on Cognitive Communications and Networking*, Vol. 6, No. 4, pp. 1146-1154, 2020.

[17] B. Li, Y. Wu, J. Song, R. Lu, T. Li, and L. Zhao, "DeepFed: Federated deep learning for intrusion detection in industrial cyber–physical systems", *IEEE Trans. Industr. Inform.*, Vol. 17, No. 8, pp. 5615–5624, 2021.

[18] Y. Qu, S. R. Pokhrel, S. Garg, L. Gao, and Y. Xiang, "A blockchained federated learning framework for cognitive computing in industry 4.0 networks", *IEEE Trans. Industr. Inform.*, Vol. 17, No. 4, pp. 2964–2973, 2021.

[19] T. T. Huong et al., "Detecting cyberattacks using anomaly detection in industrial control systems: A Federated Learning approach", *Comput. Ind.*, Vol. 132, No. 103509, p. 103509, 2021.

[20] S. Liang and H. Wu, "Edge YOLO: Real-time intelligent object detection system based on edge-cloud cooperation in autonomous vehicles", *IEEE Transactions on Intelligent Transportation Systems*, Vol. 23, No. 12, pp. 25345-25360, 2022, doi: 10.1109/TITS.2022.3158253.

[21] S. Y. Nikouei, Y. Chen, S. Song, R. Xu, B.-Y. Choi, and T. R. Faughnan, "Real-time human detection as an edge service enabled by a lightweight CNN", In: *Proc. of 2018 IEEE International Conference on Edge Computing (EDGE)*, 2018.

[22] Z. Chen, N. Lv, P. Liu, Y. Fang, K. Chen, and W. Pan, "Intrusion detection for wireless edge networks based on federated learning", *IEEE Access*, Vol. 8, pp. 217463–217472, 2020.

[23] Li, Y. Guo, D. Liu, Y. Ren, R. Hu, and Z. Guan, “Client-edge-cloud hierarchical federated learning based on generative adversarial networks”, In: *Proc. of 2023 IEEE International Conference on Knowledge Graph (ICKG)*, 2023.

[24] S. Wang, “Research towards Yolo-series algorithms: Comparison and analysis of object detection models for real-time UAV applications”, *J. Phys. Conf. Ser.*, Vol. 1948, No. 1, p. 012021, 2021.

[25] T. Diwan, G. Anirudh, and J. V. Tembhurne, “Object detection using YOLO: challenges, architectural successors, datasets and applications”, *Multimed. Tools Appl.*, Vol. 82, No. 6, pp. 9243–9275, 2023.

[26] “Avenue Dataset,” Edu.hk. [Online]. Available: https://www.cse.cuhk.edu.hk/leojia/projects/detectabnormal/dataset.html. [Accessed: Mar-2024]. S.

[27] S. Hasija, “UCF Crime Dataset.” [Online]. Available: https://www.kaggle.com/datasets/odins0n/ucf-crime-dataset/data. [Accessed: Mar-2024].

[28] “Crowd 11 Dataset.” [Online]. Available: https://paperswithcode.com/dataset/crowd11. [Accessed: Mar-2024].

[29] “Shanghaitech vision and intelligent perception(SVIP) LAB,” Github.io. [Online]. Available: https://svip-lab.github.io/dataset/campus_dataset.html. [Accessed: Mar-2024].

[30] S. Savian, M. Elahi, and T. Tillo, “Optical flow estimation with deep learning, a survey on recent advances”, In: *Deep Biometrics,* Cham: Springer International Publishing, 2020, pp. 257–287.