J. COMBIN. MATH. COMBIN. COMPUT. 127a (2025) 8361--8377

Journal of Combinatorial Mathematics and Combinatorial Computing
www.combinatorialpress.com/jcmcc

# Research on real-time monitoring and response method of AI emergency safety scenarios in energy industry based on machine vision

Sifa Qian[1], Kehong Li[2,✉]

[1] *Anhui Wanbei Coal and Electricity Group Co., Ltd., Suzhou, Anhui, 234000, China*

[2] *Yunding Technology Co., Ltd., Jinan, Shandong, 250000, China*

ABSTRACT

The study applies machine vision technology to the production and operation process of energy enterprises, and constructs a fire detection model based on improved YOLOv4 from the real-time monitoring of fire emergency safety scenarios. Based on the original YOLOv4 algorithm, the model lightens the feature extraction and feature fusion networks, and introduces CA attention mechanism in the bottom layer of the feature extraction network to improve the accuracy of target detection. An intelligent fire alarm system is built on this basis as a response method for emergency security scenarios. Comparison with the basic YOLOv4 algorithm reveals that the improved YOLOv4 algorithm reduces the parameter amount by 45.97%, improves the FPS by 27.75, and improves the mAP value by 14.10%, which achieves a better detection accuracy on the basis of greatly reducing the amount of computation and parameter count, and also achieves a better Loss value and mAP in the comparison with other detection methods. Intelligent Fire Alarm The system integrates intelligent detection, intelligent alarm, intelligent alarm receiving and intelligent alarm dispatching, and can complete the fire alarm process within 6s. In summary, it shows that the method proposed in this paper can be used in real-time monitoring of emergency security scenarios and can provide timely warning at the early stage of security hazards.

*Keywords:* YOLOv4; feature extraction; attention mechanism; machine vision; fire detection; real-time monitoring

## 1. Introduction

As an important pillar of social and economic development, the energy industry covers the fields of petroleum, mining, electric power, renewable energy, energy technology services, etc., and its safety

✉ Corresponding author.
*E-mail address:* 13361075351@163.com (K. Li).
Received 03 March 2024; Revised 10 May 2024; Accepted 25 November 2024; Published Online 15 April 2025.
DOI: 10.61091/jcmcc127a-461

production is related to the stability of the national economy and the tranquility of people's lives [1-4]. However, the energy industry faces a variety of potential risks and safety hazards, such as fires, accidents, leaks, etc., which may cause serious casualties and property losses in case of occurrence. With the changes in the world energy pattern, the energy development to low-carbon, decentralized, intelligent, automated transformation [5-7]. The demand shift in the energy consumption service market is forcing the production, storage and transportation links to be safer and more efficient, so it is necessary to rely on digital technology to improve the level of intelligent management of the energy production process. Then, it is necessary to develop a complete energy industry safety production emergency plan to respond to emergencies and minimize losses, and artificial intelligence (AI) plays a good role.

First of all, A1 technology has a strong data processing ability, which can quickly collect, organize and analyze all kinds of data information [8]. This can be more accurate for the energy industry to grasp and analyze the situation of the production process, the first time to ensure the safety of people and property, and improve its understanding of the event and decision-making level. Secondly, AI technology can help the monitoring system realize automated warning and prevent accidents as much as possible by judging abnormalities and troubleshooting and identifying potential risks [9-10]. Finally, once an emergency occurs, AI technology can immediately make an effective response and treatment, and by analyzing a large amount of historical data and real-time information, it can provide intelligent auxiliary scheduling suggestions for the emergency department, so as to optimize the allocation, emergency resource deployment and other work [11-12]. However, there are some challenges of AI technology in emergency monitoring and response. The first and foremost is the privacy issue. Data analysis using A1 technology involves a large amount of internal information [13-15]. With today's network technology, there may be malicious attacks thus stealing secrets, disrupting the industry, and even threatening the economic stability of the country. Therefore, privacy protection is imperative and cyber security cannot be ignored. In addition, A1 algorithms may have certain biases and misjudgments when dealing with complex situations [16]. In order to avoid the algorithm from being unfair to certain groups or producing wrong results, it is necessary to train the algorithm with diverse datasets and conduct systematic supervision and evaluation.

Artificial intelligence in natural disaster management mainly focuses on prevention and mitigation of injuries and preparation of plans first, followed by response measures and post-disaster recovery [17]. In contrast, literature [18] designed a real-time decision-making platform with a network of wireless low-consumption renewable energy sensing units to cope with the problem of facility structural management failures in natural disasters, realizing an almost all-encompassing emergency safety management system for monitoring, diagnosis, maintenance, escape routes, and post-disaster assessment. In addition, literature [19] utilizes a combination of three types of sensors: vibration, respiration and motion, and light detection and ranging to collect data on human activities in factories, and uses machine learning algorithms in AI technology to construct models for training and testing to achieve real-time monitoring of mechanical injuries or poisonings of factory personnel when they come into contact with equipment or materials.

Mining industry as a peak safety risk area in the energy industry, literature [20] utilized AI and cloud computing to design a virtual platform for coal mine safety training to prevent safety accidents in advance. Natural disasters are unavoidable when they occur, but the extent of the disaster is too large and may cause unpredictable loss of life and property. The oil and gas industry is also an energy industry that is prone to safety problems. Literature [21] integrates AI into HSE (Health, Safety and Environment) management systems to achieve safety in industry operations by monitoring hazardous situations in real time, automating daily inspections, optimizing management, and predicting potential risks. Among them, there is a corrosion risk factor in the petroleum sector, and literature [22] utilizes AI-driven monitoring of petroleum corrosion and constructs a corrosion prediction model to achieve automatic identification and mitigation of corrosion. In addition, the energy sector is also involved in the field of new energy vehicles, while the literature [23] utilizes AI and IoT to build an intelligent accident detection and rescue system that can simulate human thinking. The power sector is also in the energy industry, while the safety of the power grid concerns the stability of all human beings and the economy. Of course, the energy power sector, which is also the most exposed area, AI can monitor power system operations in real time, balance power supply and demand, minimize energy consumption, and optimize the management of decentralized grids in the excess of renewable energy sources [24]. And the literature [25] utilizes a combination of AI and AR technology applied to emergency training in the power grid for satellite communication equipment to achieve the training phase in the emergency plan.

In this paper, the architecture of real-time monitoring system for emergency safety scenarios is sorted out to design a detection model that meets the real-time monitoring of hazardous sources in the production process of the energy industry. YOLOv4 is used as the base model to construct the fire detection algorithm, the feature extraction network and feature fusion network of the original YOLOv4 are lightweighted, and the standard convolution in the SPP and PANet modules is improved by using the idea of depth-separable convolution. After each depth-separable convolution in the underlying feature extraction network, the Coordinate Attention attention mechanism is introduced and augmented with Mosiac data, enabling the whole network model to focus more on the flame and smoke regions. Comparison experiments on the improved YOLOv4 algorithm are carried out on the self-built dataset to verify the effectiveness of the improved actions in this paper, as well as the fire detection effect of the constructed model. Based on this, an intelligent fire alarm system based on Nodejs and Vue with front-end and back-end separation is designed to integrate the improved YOLOv4 fire detection algorithm into the alarm system, and simulation experiments are conducted to verify the superiority of the intelligent fire alarm system.

## 2. AI emergency security scenario real-time monitoring system

Restricted by the conditions of resource existence, complexity of development and utilization process, technology and equipment level and other factors, many energy industry enterprises have imperfect construction of safety monitoring system, the lack of comprehensive safety monitoring system and other problems, resulting in monitoring data is not comprehensive and accurate, it is difficult to detect potential safety hazards in a timely manner, and some of the enterprises have untimely transmission of early warning information, unclear responsibility, and irregular processing of early warning information. Some enterprises have problems such as untimely transmission of early warning information, unclear responsibility and irregular processing of early warning information.

In recent years, with the rapid development of science and technology, neural network technology has been gradually applied to energy production, and deep learning technology based on machine vision has significant effects in real-time monitoring of emergency safety scenes. In the identification of hazardous sources, convolutional neural networks perform well in image recognition. By training the neural network to use a large number of labeled images, the network is able to learn the features of the hazardous sources, enabling it to accurately detect them. For video analysis, the dynamic changes in hazard sources can be analyzed in conjunction with the surveillance video streams that are everywhere in the plant. This is very effective for monitoring safety scenarios in real time in environments such as surveillance cameras and fire alarm systems. In terms of implementation decision making, neural networks are able to process large amounts of data quickly to enable real-time decision making. This is particularly important for mission-critical tasks that require a rapid response in the early stages of a hazard.

The AI emergency safety scene real-time monitoring system needs to collect real-time video data from the monitoring area during operation and ensure smooth transmission between it and the detection system. At the same time, the system must have the ability to detect real-time video, be able to instantly display the detection results and trigger an alarm to ensure the effectiveness of the warning function. In the design of the safety hazard detection system, it is necessary to focus on its real-time, stability and reliability to ensure that early safety hazards can be detected in a timely manner during operation. The architecture of the real-time monitoring system for emergency safety scenarios is shown in Figure 1, which is mainly divided into three parts: the hardware layer, the software layer and the working layer, as explained below:

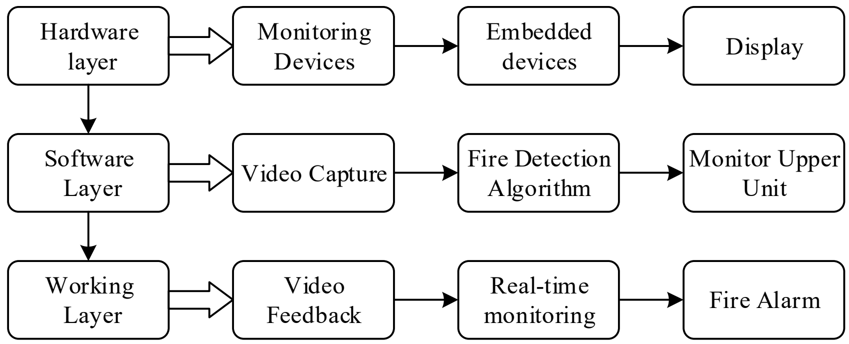


**Fig. 1.** The architecture of real-time monitoring system for emergency security scenarios

At the hardware level, it mainly includes various types of video surveillance equipment and so on. Firstly, monitoring equipment such as ordinary cameras, infrared cameras, etc. are needed, which are installed in each area that needs to be detected. Secondly, through the embedded device will be hidden security detection network and hardware equipment, and finally through the monitoring area of the monitoring screen for hidden security warning.

At the software level, it mainly includes three parts: video acquisition, safety hazard detection algorithm and monitoring host computer. Firstly, we need to collect real-time monitoring video and convert the video frame extraction into pictures. Secondly, we design a real-time monitoring algorithm to detect potential safety hazards, send the processed monitoring pictures into the safety hazard detection network for detection, and finally reflect the detection results in real-time in the monitoring host module.

At the working level, it mainly includes three parts: video feedback, real-time monitoring and hazard alert. Firstly, it is necessary to obtain the actual monitoring video information, real-time monitoring on the safety hazard detection algorithm in the software layer, and alarm if there is a fire. When the monitor receives the alarm, it will determine the location of the hazardous source through the monitoring video and rush to the scene in time to deal with it, thus preventing the hazardous source from spreading and avoiding casualties and property losses.

In summary, the design of a detection algorithm that meets the requirements of real-time monitoring and accuracy of the hazardous source is the primary task of the detection system, so this paper starts from the fire monitoring in the emergency safety scenarios in the energy industry, selects the baseline model of the fire detection algorithm, and proposes the fire detection algorithm for the real-time monitoring of the emergency safety scenarios.

## 3. Fire detection algorithm based on improved YOLOv4

As real-time is required as the primary technical goal when fire detection is performed in the energy production process, target detection algorithms often require a large amount of computation and parameter counts due to the depth of the neural network, which is difficult to meet the real-time detection of emergency safety scenarios. Considering the arithmetic power of the on-board computers, this chapter improves the YOLOv4 algorithm on the basis of it, and constructs a machine vision-based fire detection algorithm that is used for real-time monitoring and warning of emergency safety scenes.

### *3.1. YOLOv4 algorithm*

YOLO creatively treats the object detection task directly as a regression problem, combining the two phases of candidate area and detection into one. YOLOv4, as a generalized target detection model, generates some problems when performing the detection task, such as large number of parameters and computation, and difficulty in detecting small targets, etc. The YOLOv4 algorithm is an optimization of the YOLOv3 algorithm, and the main structural features of YOLOv3 algorithm are as follows:

3.1.1. Inputs. YOLOv4 has been optimized on the training input side to achieve excellent performance even on a single GPU. One of the key improvements is the adoption of Mosaic data enhancement, a method that stitches together four images, randomly scaled and cropped, into a new image to be used as training data. This approach effectively improves the generalization ability and robustness of the model, thus improving the training results on a single GPU.

3.1.2. Backbone network. YOLOv4 chose to use the CSPDarkNet53 network for feature extraction. In order to solve the problem of computationally intensive DarkNet53, YOLOv4 borrowed the structure of CSPNet (Cross Stage Partial Network), which led to the derivation of CSPDarkNet53.

3.1.3. Neck network. In YOLOv4, the model continues to follow the multi-scale output strategy of the v3 version. In order to further enhance the feature extraction capability of the model, YOLOv4 introduces the Spatial Pyramid Pooling (SPP) structure, a structure that effectively expands the sensory field of the model. Meanwhile, to better fuse multi-scale features, YOLOv4 also adopts the path aggregation network (PAN) structure. Together, these improvements enhance the performance of YOLOv4 on the target detection task.

In the SPP module, the maximum pooling of $k=\{1\times1, 5\times5, 9\times9, 13\times13\}$ is used, and then feature maps from different scales are concatenated together as outputs by performing the Concat operation to concatenate feature maps from different kernel-size pooled feature maps.YOLOv4 introduces a bottom-up feature pyramid after the FPN layer. This combination makes the neck structure not only able to deliver semantic features from top-down, but also able to deliver precise localization features from bottom-up, thus improving the accuracy and efficiency of target detection.

3.1.4. Loss function. In order to optimize the detection effect of the model, YOLOv4 chooses CIOU as the loss function. Compared with IoU and GIoU, CIOU not only considers the overlap rate between target and anchor, but also integrates factors such as the distance between them, scale difference, and specific penalty terms. This comprehensive consideration makes the regression process of the target box more stable and effectively avoids problems such as dispersion that may occur during the training process, thus improving the detection performance of the model. The formula of CIOU is shown in equation (1):

$$\begin{cases} v = \dfrac{4}{\pi^2}\left(\arctan\dfrac{w^{s'}}{h^{v}} - \arctan\dfrac{w}{h}\right)^2 \\ IoU = \left|\dfrac{B \cap B^{s}}{B \cup B^{v}}\right| \\ \beta = \dfrac{v}{1 - IoU + v} \\ CIoU = IoU - \left(\dfrac{\rho^2\left(b, b^{s^s}\right)}{C^2} + \beta v\right) \\ L_{Clov} = 1 - IoU + \dfrac{\rho^2\left(b, b^{vz}\right)}{C^2} + \beta v \end{cases} \tag{1}$$

Where $w^{gt}$ and $h^{gt}$ are the width and height of the real frame, $w$ and $h$ are the width and height of the predicted frame, $B$ is the predicted frame, $B^{gt}$ is the real frame, $c$ is the diagonal distance between the real frame and the outer rectangle of the predicted frame, and $\rho$ is the Euclidean distance between the coordinates of the center points of the real frame and the predicted frame.

3.1.5. DropBlock regularization. YOLOv4 introduces the DropBlock regularization technique, whose main purpose is to overcome the limitations of the traditional Dropout method of randomly discarding features.The DropBlock technique reduces the model's overdependence on the training data by randomly discarding features in the adjacent correlation regions of the neural network, thus reducing the risk of overfitting. Specifically, in each training iteration, DropBlock introduces the concept of learning some of the network weights and compensating for the weight matrix to reduce the dependencies between neurons, increasing the robustness of the network and reducing overfitting. The DropBlock technique can be implemented by adding a DropBlock layer between the layers of the neural network, which randomly discards a certain percentage of the features in the neighboring regions of interest, thus making the network generalizable and robust.

3.1.6. CmBN. Cross mini-Batch Normalization (CmBN) is an improvement on Batch Normalization (BN), which is a special layer in neural networks that reduces gradient vanishing by normalizing the data in each batch, and the addition of a BN layer to the network can speed up convergence and improve model Accuracy. In BN, the mean and variance are calculated for each batch of data, and then the BN operation is performed on it, which is expressed by the formula as shown in equation (2):

$$\hat{x}_i = \frac{x_i - \mu_i}{\sqrt{\delta^2 + \varepsilon}} \tag{2}$$

$$y_i = \gamma \hat{x}_i + \beta \tag{3}$$

Where $x_i$ is the input of the $i$ nd sample with dimension $n$, $\mu_i$ and $\delta_i^2$ are the mean and variance of all the samples in that dimension, and $\hat{x}_i$ is the input of the $i$ th after the change of BN. $\varepsilon$ is a very small value to prevent the denominator from being 0. BN will introduce two learnable parameters $\gamma$ and $\beta$ for scaling and flattening the normalized values, and the final output obtained is shown in Equation (3).BN calculates the mean and variance on each batch, and a batch that is too large or too small will have an effect on the training, and in order to overcome this problem, CmBN is proposed, which will not only learn the mean and variance, it will also adjust the features on each channel differently according to the input condition information, so that the network can learn different features. Specifically, suppose the input features are $X \in R^{C \times H \times W}$, $C$ is the number of channels, $H$ and $W$ are the height and width of the features, respectively, and for CmBN, the input conditional information is $Z \in R^{C_2}$, where $C_2$ is the dimension of the conditional information. Firstly, the mean and variance on the channels are calculated respectively, as shown in Eqs. (4) and (5):

$$\mu_c = \frac{1}{H \times W} \sum_{k=1}^{H} \sum_{w=1}^{W} X_{c,h,w} \tag{4}$$

$$\delta_c^2 = \frac{1}{H \times W} \sum_{k=1}^{H} \sum_{w=1}^{W} \left( X_{c,h,w} - \mu_c \right)^2 \tag{5}$$

Normalize the features for each channel:

$$\hat{X}_{c,h,w} = \frac{X_{c,h,w} - \mu_c}{\sqrt{\delta_c^2 + \varepsilon}} \tag{6}$$

Two learnable parameters $\gamma_c$ and $\beta_c$ are then introduced to smooth the normalized features:

$$\hat{X}_{c,h,w} = \gamma_c \hat{X}_{c,h,w} + \beta_c \tag{7}$$

Finally, the normalized features are conditioned according to the input conditional information Z to obtain the final output:

$$Y_{c,h,w} = Z_c \hat{X}_{c,h,w} \tag{8}$$

### *3.2. Improvement of the YOLOv4 algorithm*

According to the characteristics of fire detection in emergency safety scenarios, the network structure of YOLOv4 is still too large for embedded platforms. Therefore, the YOLOv4 algorithm is firstly designed to be lightweight to reduce the number of parameters and computation of the on-board computer, and the feature extraction network and feature network are improved. Then for the problem of low accuracy of smoke and flame detection, the Coordinate Attention attention mechanism is introduced, which makes the network extract more features from the region near the target object, thus improving the detection accuracy.

3.2.1. Improvement of feature extraction network. Deeply separable convolutional networks are used instead of CSPDarknet-53 to achieve model acceleration by greatly reducing the number of parameters and computation without degrading the model detection accuracy too much.

1) Backbone network improvement. Deeply separable convolutional network (MobileNet_V1) provides lightweight functions for mobile and embedded devices, and the main idea is to use deeply separable convolution to replace the traditional standard convolution. Deeply separable network separates channels and space, firstly the input image is deeply convolved, and the generated result is convolved point by point for feature extraction on all channels of the feature map. Where the convolution kernel for deep convolution is the same as the number of channels in the input image, deep convolution does not change the number of channels in the input feature map, it only changes the size of the feature map. The convolution kernel of point-by-point convolution is 1×1 in size, which convolves all channels of the input feature map, and the number of its convolution kernels determines the number of channels of the output feature map, but does not change the size of the feature map.

Assuming a feature map with input feature map size $D_{in} \times D_{in} \times M$, where $M$ is the number of channels in the input feature map, the final output feature map is required to be $D_{out} \times D_{out} \times N$, where the number of channels in the final output feature map is $N$.

For standard convolution processing, let the standard convolution kernel be $D_k \times D_k \times M$, $Padding = 0$, $Stride = 1$, and $N$ processing is required, then there are $D_{out}$ computational formulas, the total number of parameters $P_C$ for standard convolution, and the total computational $C_C$ formulas are shown in Eq. (9), Eq. (10), and Eq. (11):

$$D_{out} = D_{in} - D_k + 1 \tag{9}$$

$$P_C = D_k \times D_k \times M \times N \tag{10}$$

$$C_C = D_k \times D_k \times M \times N \times D_{out} \times D_{out} \tag{11}$$

For depth separable convolution is divided into two steps: the first step of depth convolution, let the size of the depth convolution kernel is $D_k \times D_k \times M$, $Padding = 0$, $Stride = 1$, then the depth convolution part of the parametric quantities $P_D$, the computational amount of $C_D$ formula as shown in Equation (12), Equation (13):

$$P_D = D_k \times D_k \times M \tag{12}$$

$$C_D = D_k \times D_k \times M \times D_{out} \times D_{out} \tag{13}$$

The second step of point-by-point convolution, set the size of the point-by-point convolution kernel as $1 \times 1 \times M$, $Padding = 0$, $Stride = 1$, then the number of parameters in the point-by-point convolution part of the number of $P_P$, the amount of computation of $C_P$ formula as shown in Equation (14), Equation (15):

$$P_P = M \times N \tag{14}$$

$$C_P = D_{out} \times D_{out} \times M \times N \tag{15}$$

When the input and output dimensions are the same, the total number of parameters $P_{DW}$ and the computational amount $C_{DW}$ of the depth separable convolution are calculated as shown in Eqs. (16) and (17):

$$P_{DW} = P_D + P_P = D_k \times D_k \times M + M \times N \tag{16}$$

$$\begin{aligned} C_{DW} &= C_D + C_P = D_k \times D_k \times M \times D_{out} \times D_{out} \\ &\quad + D_{out} \times D_{out} \times M \times N \end{aligned} \tag{17}$$

Therefore, the standard convolution and depth separable convolution parametric quantities, computational logarithms such as Eqs. (18) and (19):

$$\frac{P_{DW}}{P_C} = \frac{D_k \times D_k \times M + M \times N}{D_k \times D_k \times M \times N} = \frac{1}{N} + \frac{1}{D_k^2} \tag{18}$$

$$\begin{aligned} \frac{C_{DW}}{C_C} &= \frac{D_k \times D_k \times M \times D_{out} \times D_{out} + D_{out} \times D_{out} \times M \times N}{D_k \times D_k \times M \times D_{out} \times D_{out} \times N} \\ &= \frac{1}{N} + \frac{1}{D_k^2} \end{aligned} \tag{19}$$

From Eq. (18) and Eq. (19), it can be seen that in the case of the general standard convolution kernel of 3, the number of parameters of the model and the computation of the model can be reduced by about 8 times with the use of the depth separable convolution, which realizes the purpose of the lightweight YOLOv4 network, and greatly improves the efficiency of the model's detection and the network's computational speed.

2) Feature fusion network improvement. Based on the feature extraction network in YOLOv4, the same idea of depth separable convolution is taken, and all the 3*3 convolutions in the SPP module, FPN+PANet module are replaced with depth separable convolutions, and the optimized modules are denoted as SPP_N, PANet_N. The number of parameters and computation of the whole model are further reduced.

3.2.2. Optimization of attention mechanisms. After the network model is lightweighted, in order to further improve the accuracy of the model in detecting the two objects in the dataset, flame and smoke, it is proposed to introduce an attention mechanism in the bottom layer of the feature extraction network. In order to better extract the feature information of flame and smoke in the input fire image, the Coordinate Attention mechanism is added to the convolution operation of MobileNet_V1 to suppress the response of background information features outside the fire region.

Coordinate Attention encodes the channel relationship and long-term dependency through precise position information, and the specific operation is divided into two steps: Coordinate information embedding and Coordinate Attention generation.

1) Coordinate information embedding. Traditional channel attention uses a global pooling operation for global encoding, but this makes all the information spatially compressed, so in order to be able to capture accurate positional information, the global pooling first needs to be decomposed and transformed into a pair of one-dimensional feature encoding operations, which, for a given input $X$, encodes each channel along the horizontal and vertical coordinates, respectively, using pooling dimensions $(H,1)$ and $(1,W)$ along the two spatial directions to aggregate features and obtain a pair of direction-aware feature maps. This captures the long-term dependencies between channels and preserves the precise location information along the spatial directions.

2) Coordinate Attention Generation. After passing the transformations in the information embedding, the transformations from the first step are stacked and transformed using the convolution operation. The Coordinate Attention attention mechanism is introduced into each depth-separable convolutional module of the feature extraction network MolbilenetV1_N. Where the Coordinate Attention module is denoted as CA, adding the CA attention mechanism in the form of a residual module after the depth separable convolution enables better feature extraction.

### *3.3. Experimental results and analysis*

3.3.1. Data sets and related configurations. In order to fully utilize the algorithmic model of this experiment and achieve a certain training effect, at the same time, in order to obtain more diversified and rich data samples in different scene information, we find clearer related images on the Internet and record the fireworks video of the real scene, and obtain the target image of the fireworks by skipping frame interception of the video. In order to improve the robustness of the framework model, a total of 4,000 images were collected for this investigation, which were all collected by ourselves and characterized by different backgrounds and different sizes and forms of fires, and were divided into forest scenes, factory scenes, and as a comparison scenes that did not contain fires in the ratio of 7:2:1. The dataset was divided into dataset, validation set, and test set in the form of 7:2:1. train_num=2800, val_num=800, test_num=400.

3.3.2. Optimization module validation. The network setup is as follows: first input a 512×512 matrix into YOLOv4, train and validate the original model set to 150 epochs, train each epoch 400 times for a total of 60,000 iterations. The batch size Bacthsize was set to 64, the learning rate was set to 0.0001, and the hyperparameters with the best detection were selected after debugging the model several times. This dataset was then used to train the improved YOLOv4 model and the model with lightweight processing, using the same training methodology and 100 model parameters with the smallest validation loss after training. The optimizer used the Adam method and the loss function was chosen as cross-entropy loss.

The experiment firstly, SSD, Faster-RCNN, YOLOv3, YOLOv4, and CenterNet are detected and compared to the dataset respectively, and the results of the experimental comparison are shown in Fig. 2. According to the experimental results, the combined precision, recall, and mAP are verified, and the AP value and recall of YOLOv4 are 74.71% and 66.58%, respectively, which are higher than the verification results of other algorithms, and are more suitable to be used as the base model of this experiment.

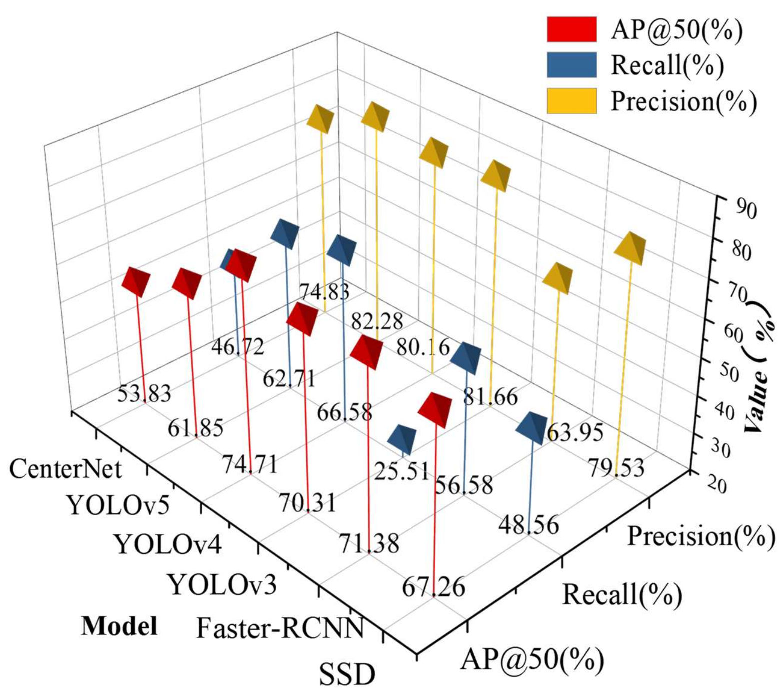


**Fig. 2.** Comparison results of the experiment

Second, MobileNet_V1 is added to the YOLOv4 network to enhance the feature extraction of the network model, and Coordinate Attention (CA) is fused to focus the attention to the range where there are fires. The detection results of different algorithms are shown in Table 1, and the average accuracy of the typical YOLOv4 model for detecting fires is 61.75%, but the accuracy of the original model is low due to the different shapes of flames, hue backgrounds, and inconsistent manual labeling data. Adding MobileNet_V1 is significant, the amount of parameters is reduced by 43.58%, and the accuracy rate reaches 77.83% after adding MobileNet_V1, which is 10.38% higher than the original model. With Coordinate Attention processing, the accuracy of the improved YOLOv4 model is increased by 3.72%, while the number of parameters is reduced by about 6.41MB.

**Table 1.** Test results of different algorithms

| | YOLOv4 | YOLOv4+ MobileNet_V1 | YOLOv4+ MobileNet_V1+CA |
|---|---|---|---|
| AP@50 (%) | 61.75 | 65.12 | 74.72 |
| Recall (%) | 62.12 | 67.27 | 73.66 |
| Precision (%) | 67.45 | 77.83 | 81.55 |
| FPS | 36.93 | 45.31 | 64.68 |
| Parameter/MB | 266.46 | 150.37 | 143.96 |

3.3.3. Training and analysis. Four models, YOLOv4, YOLOv4-Tiny, YOLOv3-Tiny and the improved YOLOv4 in this paper are trained respectively, and the learning rate of the modified model is 0.000232 after the training reaches 25,000 times, and the Loss value of the modified model can converge to a smaller stable value. The training log is saved to a document, and the Loss parameter during the training process is extracted and plotted after the training is completed.

The Loss curves of the four models trained after 30,000 times of training are shown in Figure 3. Among the four models, the Loss value of the improved model in this paper decreases the fastest, and the Loss value reaches 0.11 after 5000 times of training, and the Loss tends to stabilize, and finally stays around 0.04. The Loss value of the YOLOv4-Tiny model decreases slower, and the final Loss value stabilizes around 0.10, which is higher than that of the improved model. The YOLOv4 model has a slow process of decreasing Loss value due to the overly complex structure, the high number of network layers and the large number of parameters, etc., which finally stabilizes at around 0.12. The Loss value of the stabilized YOLOv3-Tiny is higher than the other three models, which is around 0.32. From the results, it can be seen that the improved YOLOv4 model has a better training effect.

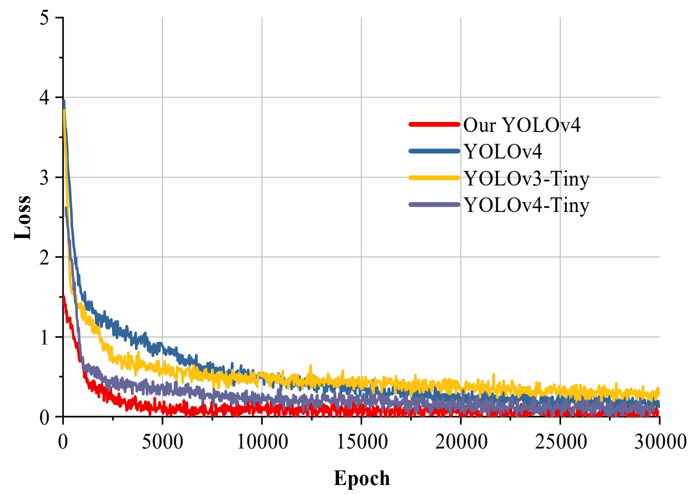


**Fig. 3.** The Loss curve of four model training

The current metrics used to evaluate the detection effectiveness of the YOLO model are the average precision (AP) of single class target detection and the mean of the average precision of all class target detection (mAP), with larger values of AP and m AP indicating better results. Where AP denotes the sum of each type's recall per prediction multiplied by the accuracy corresponding to the current recall, and mAP denotes the mean of all APs. The four trained models are tested separately using the test dataset, and the mAP values of the four model trainings are shown in Fig. 4.

The mAP of the improved YOLOv4 model in this paper is 93.56%, YOLOv4-Tiny is 87.41%, YOLOv4 is 30.25%, and YOLOv3-Tiny is 85.82%, and the improved YOLOv4 model is improved compared to the other three models by 6.15%, 63.31% and 7.74 percent. As can be seen in Figures 3 and 4, the

YOLOv4 model has a lower Loss and also the lowest mAP, which is an overfitting phenomenon due to the fact that the model is too large and the training dataset is too small.

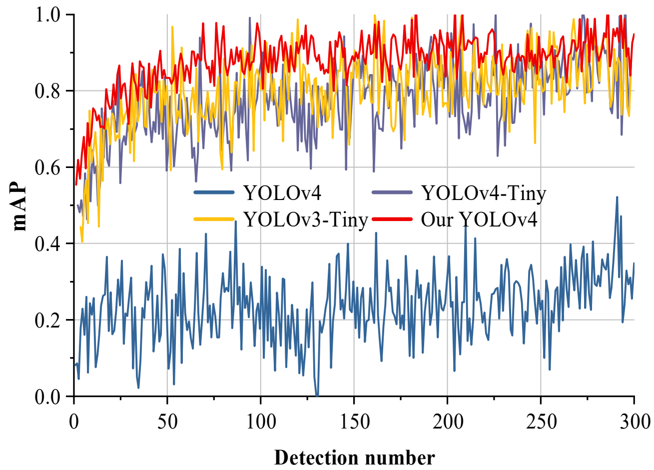


**Fig. 4.** The mAP value of four model training

Using the models to detect the images of the forest scene and the factory scene respectively, the average detection accuracies of the four models for the two types of flames are shown in Fig. 5, which shows that the detection accuracies of the improved YOLOv4 model for each type of flames are 89.27% and 92.98%, which are 4.08% and 2.70% higher than the YOLOv4-Tiny, and 14.76% and 3.56% higher than the YOLOv4 74.73% and 32.90%, and 14.76% and 3.56% than YOLOv3-Tiny, indicating that the improved YOLOv4 model has the best detection effect among the four models tested in comparison.

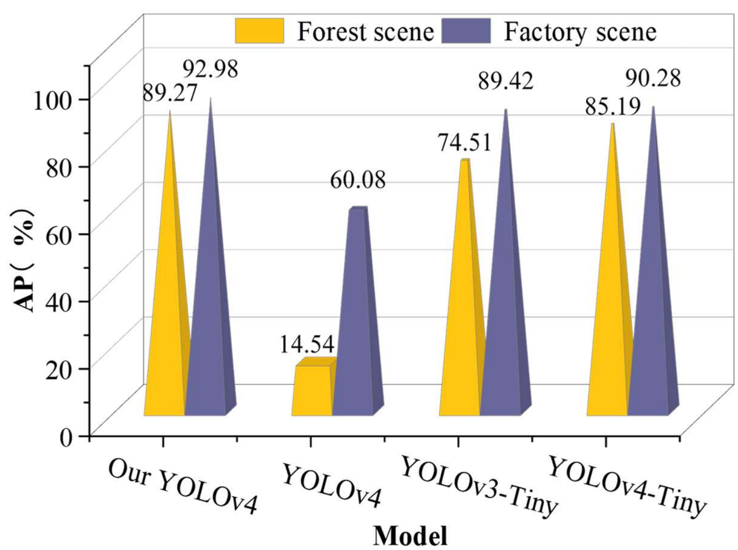


**Fig. 5.** The average detection accuracy of various targets of four model

In order to evaluate the performance of the four models, the models are tested on the GPU platform. The four models are run on the GPU terminal to detect video images with a resolution of 384×288, and the results of the comparative test of the detection performance of each model are shown in Fig. 6. The detection time per frame of the four models is 0.029s, 0.203s, 0.019s and 0.026s, respectively, among which YOLOv4 takes the longest time, YOLOv3-Tiny takes the shortest time, and the time taken by the improved YOLOv4 model is 0.01s more than that of the YOLOv3-Tiny model, and although the time taken by the improved YOLOv4 model is 0.01s more than that of YOLOv3-Tiny, it is still able to detect video images of 384×288 resolution. Tiny by a small amount, but it can still meet the requirements of practical applications.

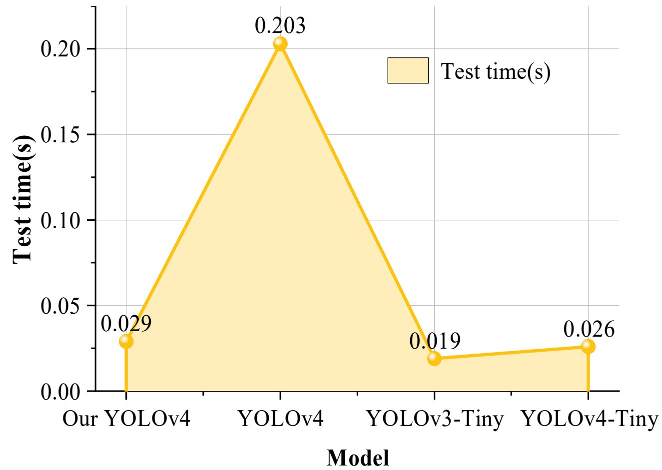


**Fig. 6.** The performance comparison test results of each model

## 4. Response methodology for AI emergency security scenarios

Model deployment and system implementation is the final phase of a deep learning project and is a way to integrate a deep learning model into an existing production environment in order to start using it to make actual business decisions based on data. This chapter carries out the related intelligence module focusing on the realization of the model's implementation and the design of a response methodology for AI emergency safety scenarios in the energy industry.

### *4.1. Intelligent Fire Alarm System*

Cross-platform fire detection is implemented using the improved YOLOv4 model described in the previous section, including fire desktop detection software, Android client application software and web pages to visualize fire detection.

Implement basic fire local detection software using PyQt5 and import the trained improved YOLOv4 model weights file into the project. Combine with TorchScript to develop Android application that can realize fire detection. Flask and OpenCV are used to realize viewing real-time video streams in the browser and fire detection.

To improve the alarm efficiency of the fire protection system and better integrate the fire detection model, design a front-end and back-end separated intelligent fire alarm system based on Nodejs and Vue. The system combines the designed model to maximally optimize the alarm receiving process and achieve intelligent detection, intelligent alarm, intelligent alarm receiving and intelligent alarm dispatching.

The system architecture of the smart fire alarm system is shown in Figure 7.

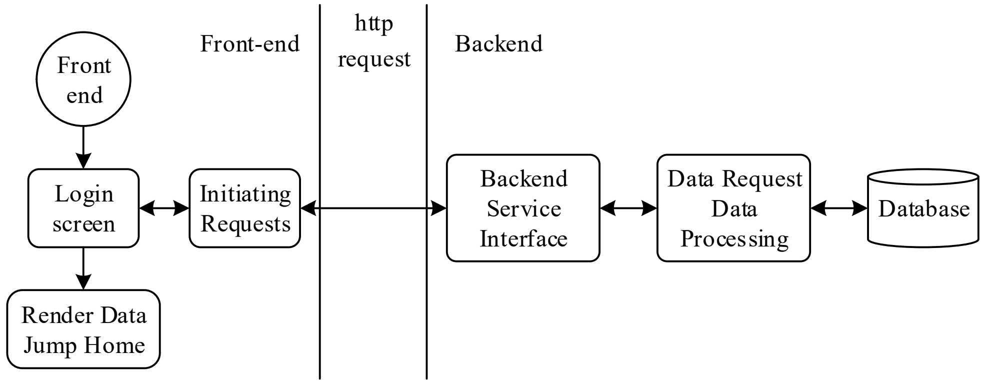


**Fig. 7.** System architecture of intelligent fire alarm system

First of all, according to the need to establish a database table, write the corresponding back-end service interface to call the data stored in the database, the front-end through the http request to establish a connection with the back-end, the back-end will get the data back to the front-end, in the interface for rendering. When the data is modified, after the display interface operation is completed, the modified data is passed to the back-end for corresponding processing and stored in the database.

After a fire occurs, the system has two ways of receiving alarms. The first one is the traditional telephone alarm, where the relevant personnel fill in the alarm form to complete the alarm work through the descriptive information of the alarmer. The second is that the system can automatically obtain the alarm request and the description of the scene information from the cross-platform automatic fire detection software and generate the form.

In the automatic alarm process, the system receives the alarm request from the improved YOLOv4 model for detecting the fire occurrence, triggers the alarm in the command center to complete the initial rapid alarm work, and the scene situation extracted from the improved YOLOv4 model is used as important information to generate the alarm form for subsequent decision-making.

Upon receiving an alarm, the system displays the coordinates of the alarm point on the map and intelligently matches the most suitable organization and vehicle to handle the firefighting task based on the location information of the existing firefighting points and other reference factors such as equipment and materials. After the fire point and the fire truck are identified, the interface of the map is called to plan the route for them and display the position of the vehicle in real time.

### *4.2. Simulation experiments*

In this paper, time is used as an indicator to demonstrate the superiority of the intelligent fire alarm system, and a simulation experiment is conducted at the local fire station to compare the traditional alarm process with the intelligent fire alarm system, respectively, and the time comparison before and after the application of the intelligent fire alarm system is shown in Figure 8. Before using the intelligent fire alarm system, ideally, it takes a total of 78s from the discovery of fire to the dispatching of vehicles, while the system designed in this paper only uses less than 6s, which buys valuable time for the fire system to deal with the fire, which also reflects the superiority of the intelligent system instead of manpower.

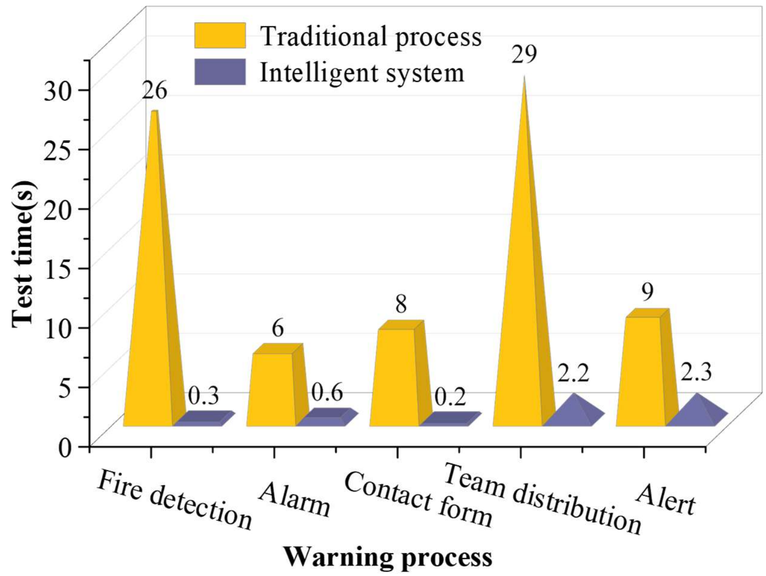


**Fig. 8.** The contrast of the time of intelligent fire alarm system

## 5. Conclusion

In this paper, based on the real-time monitoring of emergency safety scenarios in the energy industry, the feature extraction network of the YOLOv4 algorithm is improved and the attention mechanism is optimized, so as to propose a fire detection algorithm based on the improved YOLOv4. Then we build an intelligent fire alarm system and carry out simulation experiments to explore the response method of emergency safety scenarios.

The base network is selected for the experiments, and the target precision in the verification set indicates that the AP value and recall of YOLOv4 are higher than the comparison method, which are 74.71% and 66.58%, respectively, and thus it is better than YOLOv3 and YOLOv5 in terms of overall performance, and YOLOv4 is more modular compared to YOLOv5. A new model is generated based on the original model using MobileNet_V1 and Coordinate Attention, the accuracy of the model after MobileNet_V1 is added is improved by 10.38% and the number of parameters is reduced by 43.58% compared to YOLOv4, and the accuracy of the model after both are added is improved by 14.10% compared to YOLOv4. The improved YOLOv4 fire detection model is more in line with the engineering practice standards and more practical.

In the experiments with other detection algorithms, the improved YOLOv4 in this paper has more superior performance in Loss value and mAP, which are 0.04 and 93.56%, respectively. The improved model achieves a mAP of 91.13% for two different types of flames, which is 53.82%, 9.16%, and 3.39% higher compared to YOLOv4, YOLOv4-Tiny, and YOLOv3-Tiny models, respectively. In addition, the intelligent fire alarm system is able to rapidly complete the fire alarm process with an overall time spent within 6s, which is faster than the traditional process, reflecting the efficiency of the designed AI emergency safety scenario response method.

Currently the fusion intelligent warning technology applied to the energy industry is gradually developing, this study proposes a fire detection algorithm based on machine vision, and after realizing the accurate recognition of fire information in the video, the recognition technology is modularized and integrated into the video monitoring system, which can be used as an AI emergency safety scenario real-time monitoring method to detect fires early and thus avoid fires, and to safeguard the energy industry enterprises' Safe production.

## References


[1] Pareglio, S., Zoboli, R., Mazzanti, M., Bonacorsi, L., Chioatto, E., D'Amato, A., ... & Zecca, E. (2020). Energy and the circular economy: how to fill the integration gap?. Briefs.

[2] Baidya, D., Ghoreishi-Madiseh, S. A., & Scoble, M. J. (2019). Energy portfolio versus sustainability of mining industry. 9th International Conference on Sustainable Development in the Minerals Industry.

[3] Novosel, D. (2019). Energy sustainability: guiding the power & energy industry [leader's corner]. IEEE Power and Energy Magazine, 17(4), 12-14.

[4] Al-Fattah, S. (2020). The evolving role of oil and gas companies in the energy industry. SSRN Electronic Journal.

[5] Fankhauser, S., & Jotzo, F. (2017). Economic growth and development with low-carbon energy. Wiley Interdisciplinary Reviews: Climate Change.

[6] Wu, J., Zuidema, C., & Gugerell, K. (2018). Experimenting with decentralized energy governance in China: The case of New Energy Demonstration City program. Journal of Cleaner Production, 189, 830e838.

[7] Tabor, D. P., Roch, L. M., Saikin, S. K., Kreisbeck, C., Sheberla, D., Montoya, J. H., ... & Aspuru-Guzik, A. (2018). Accelerating the discovery of materials for clean energy in the era of smart automation. Nature Reviews. Materials, 3(5).

[8] Alenizi, F. A., Abbasi, S., Mohammed, A. H., & Rahmani, A. M. (2023). The artificial intelligence technologies in Industry 4.0: A taxonomy, approaches, and future directions. Computers & Industrial Engineering, 109662.

[9] Gadde, H. (2023). AI-Driven Anomaly Detection in NoSQL Databases for Enhanced Security. International Journal of Machine Learning Research in Cybersecurity and Artificial Intelligence, 14(1), 497-522.

[10] Wong, L. W., Tan, G. W. H., Ooi, K. B., Lin, B., & Dwivedi, Y. K. (2024). Artificial intelligence-driven risk management for enhancing supply chain agility: A deep-learning-based dual-stage PLS-SEM-ANN analysis. International Journal of Production Research, 62(15), 5535-5555.

[11] Challoumis, C. (2024, October). BUILDING A SUSTAINABLE ECONOMY-HOW AI CAN OPTIMIZE RESOURCE ALLOCATION. In XVI International Scientific Conference (pp. 190-224).

[12] Wang, X., Lin, X., & Shao, B. (2022). How does artificial intelligence create business agility? Evidence from chatbots. International journal of information management, 66, 102535.

[13] Mylrea, M., & Robinson, N. (2023). Artificial Intelligence (AI) Trust Framework and Maturity Model: Applying an Entropy Lens to Improve Security, Privacy, and Ethical AI. Entropy, 25(10).

[14] Dash, B., Ansari, M. F., Sharma, P., & Ali, A. (2022). Threats and opportunities with AI-based cyber security intrusion detection: a review. International Journal of Software Engineering & Applications (IJSEA), 13(5).

[15] Rahman, M. M., & Siddika, A. (2023, June). Security Risk and Attacks in AI: A Survey of Security and Privacy. In Proc. of the 17th IEEE International Workshop on Security, Trust, and Privacy for Software Applications.

[16] O'Connor, S., & Liu, H. (2024). Gender bias perpetuation and mitigation in AI technologies: challenges and opportunities. AI & SOCIETY, 39(4), 2045-2057.

[17] Sun, W., Bocchini, P., & Davison, B. D. (2020). Applications of artificial intelligence for disaster management. Natural Hazards, (3), 2631-2689.

[18] Merenda, M., Pratico, F. G., Fedele, R., Carotenuto, R., & Della Corte, F. G. (2019). A real-time decision

platform for the management of structures and infrastructures. ELECTRONICS, 8(10), 1180.

[19] Nwakanma, C. I., Islam, F. B., Maharani, M. P., Lee, J. M., & Kim, D. S. (2021). Detection and Classification of Human Activity for Emergency Response in Smart Factory Shop Floor. Appl. Sci, 11, 3662.

[20] Li, M., Sun, Z., Jiang, Z., Tan, Z., & Chen, J. (2020). A Virtual Reality Platform for Safety Training in Coal Mines with AI and Cloud Computing. Discrete Dynamics in Nature and Society, 2020.

[21] Aderamo, A. T., Olisakwe, H. C., Adebayo, Y. A., & Esiri, A. E. (2024). AI-driven HSE management systems for risk mitigation in the oil and gas industry. Comprehensive Research and Reviews in Engineering and Technology, 2(1), 1-22.

[22] Khalaf, A. H., Xiao, Y., Xu, N., Wu, B., Li, H., Lin, B., ... & Tang, J. (2024). Emerging AI technologies for corrosion monitoring in oil and gas industry: A comprehensive review. Engineering Failure Analysis, 155, 107735.

[23] Pathik, N., Gupta, R. K., Sahu, Y., Sharma, A., Masud, M., & Baz, M. (2022). AI Enabled Accident Detection and Alert System Using IoT and Deep Learning for Smart Cities. Sustainability, 14(13), 1-24.

[24] Liu, Z., Gao, Y., & Liu, B. (2022). An artificial intelligence-based electric multiple units using a smart power grid system. Energy Reports, 8, 13376-13388.

[25] Zheng, S., Zhang, M., & Zhou, H. (2021, November). Application of Augmented Reality Technology and Artificial Intelligence Satellite Communication Equipment in Power Grid Emergency Training. In Journal of Physics: Conference Series (Vol. 2074, No. 1, p. 012093). IOP Publishing.