1. Ali Abdulaziz Alawad [1], 2. Muhammad Tajammal Chughtai[2]

University of Hail, Saudi Arabia (1), University of Hail, Saudi Arabia (2)
ORCID: 1. 0000-0001-7907-9169; 2. 0000-0001-9847-6776

doi:10.15199/48.2023.12.28

IoT-based Real-time Passenger Safety System with Machine Vision at the Edge (Mez) Technology

***Abstract**. Passenger safety is a critical issue in the transportation industry. There is an additional concern for children's safety with other regular issues. And that is the risk of having accidents because of the childish act of putting hands, heads, or the upper half of the body out of the window. Children are curious and fun-loving and enjoy school bus time with their friends. Their activities are not always pleasant to adults, which easily distract drivers. That is why the stable emotional state of the drivers of school buses is essential. This paper presents an IoT-based innovative passenger safety system developed to keep the safety concerns associated with school buses. The IR-based sensor in this project prevents passengers from crossing safety limits outside the window. A well-optimized Convolutional Neural Network (CNN) has been designed and developed in this paper to predict the risk level by reading the emotional states of the driver. Real-time video transmission is essential to recognize the driver's emotional state. However, it is severely hampered by network latency. This paper incorporates the Machine Vision at the Edge (Mez) technology to solve the latency issue and effectively detect the driver's emotion in real time. This innovative safety system is a potential solution to the unaddressed safety concern of children's school buses. This paper's unique approach to solving a practical problem strengthens bus passenger safety.*

Streszczenie.** Bezpieczeństwo pasażerów jest kluczową kwestią w branży transportowej. Dodatkową troską o bezpieczeństwo dzieci są inne regularne kwestie. I to jest ryzyko wypadku na skutek dziecinnego wystawiania rąk, głów lub górnej części ciała przez okno. Dzieci są ciekawskie i kochają zabawę. Lubią spędzać czas w autobusie szkolnym z przyjaciółmi. Ich zajęcia nie zawsze są przyjemne dla dorosłych, co łatwo odwraca uwagę kierowców. Dlatego tak istotny jest stabilny stan emocjonalny kierowców autobusów szkolnych. W artykule przedstawiono innowacyjny system bezpieczeństwa pasażerów oparty na IoT, opracowany w celu spełnienia wymagań bezpieczeństwa związanych z autobusami szkolnymi. Zastosowany w tym projekcie czujnik podczerwieni zapobiega przekraczaniu przez pasażerów granic bezpieczeństwa za oknem. W tym artykule zaprojektowano i rozwinięto dobrze zoptymalizowaną konwolucyjną sieć neuronową (CNN), aby przewidywać poziom ryzyka na podstawie odczytu stanów emocjonalnych kierowcy. Transmisja wideo w czasie rzeczywistym jest niezbędna do rozpoznania stanu emocjonalnego kierowcy. Jednakże jest to poważnie utrudniane przez opóźnienia sieci. W artykule wykorzystano technologię Machine Vision at the Edge (Mez), aby rozwiązać problem opóźnień i skutecznie wykrywać emocje kierowcy w czasie rzeczywistym. Ten innowacyjny system bezpieczeństwa jest potencjalnym rozwiązaniem nierozwiązanych problemów związanych z bezpieczeństwem autobusów szkolnych dla dzieci. Unikalne podejście niniejszego artykułu do rozwiązania praktycznego problemu zwiększa bezpieczeństwo pasażerów autobusów.* (System bezpieczeństwa pasażerów w czasie rzeczywistym oparty na IoT z technologią Machine Vision at the Edge (Mez)***).

Keywords: IOT, Public Transportation, Transportation safety.
Słowa kluczowe: IOT, Transport publiczny, Bezpieczeństwo transportu.

Introduction

The bus is a popular public transportation for high capacity and low cost. Different organizations offer it as official transpiration for the employees. And it is also widely used as transport for educational institutions [1]. Buses dedicated to pick up and drop elementary school students maintain extra precautions. These precautions include cross-checking the dropping zone, assisting in crossing the road, maintaining low speed, and using caution symbols [2]. However, a few factors are out of the control of the driver or the supporting staff on the bus. Putting hands, heads, or the upper half of the body outside the window is one of them. It can cause fetal accidents at any time. Driver's attention level is another sensitive issue in passenger safety on the bus. The unstable emotional state of the driver can cause deadly accidents [3]. This paper proposes an IoT-based innovative passenger safety system that mitigates these safety risks. We have developed an IoT device consisting of Infrared (IR) sensors, a 4G LTE module, Raspberry Pi, and a Pi camera. The proposed system alerts the transportation management authority if anyone moves their body part outside the window. The alert is received by the onboard transport supporting staff as well. Frequently altering a driver causes distraction which increases the probability of accidents. It is essential to understand the emotional state of the driver as well to ensure transportation safety [4]. However, IR sensor-based systems cannot mitigate the risks of accidents caused by drivers' emotional distress. We developed an optimized Convolutional Neural Network (CNN) and trained it to recognize the driver's emotion. An innovative algorithm has been developed and presented in this paper that uses the prediction from the CNN and predicts the risk level based on the facial emotion of the driver. Real-time facial emotion recognition using Raspberry pi and sending the report to respected authorities over the internet is a computationally intensive task [5]. It also suffers from network latency issues [6]. This paper also presents an innovative solution to IoT applications that suffer from network latency issues using Machine Vision at the Edge (Mez) technology [7]. The Mez is a real-time video communication technology for latency-sensitive environments. IoT Edge systems usually suffer from high latency issues [8]. The proposed IoT device is installed in a bus that moves in different areas. As a result, getting a stable network with high bandwidth and low latency is more challenging in this context [9]. The Mez optimally maintains the trade-off between latency tolerance and video quality. Because of being resource a constraint device, it is not practical to run computationally intensive tasks like emotion recognition in IoT devices [10]. The proposed system uses Mez to adjust the video frame quality based on the dynamically concurrent latency status and transmit the frame to the cloud applications in real time. It also uses the temporary storage facility of the Mez technology to overcome the challenges of network coverage issues. It is a potential technology to enable IoT devices to tolerate up to 10x latency variations with 4.2% performance compromise [7]. This paper presents the solution to passenger safety in school buses using IoT devices, CNN, and Mez. Combining these three technologies to strengthen passenger safety is the first of its kind to our best knowledge. The core contributions of this paper have been listed below:

• Development of an innovative passenger safety system by combining IR sensor and CNN-based IoT device.
• Design and implement an optimized network for driver's emotion recognition.

• Preparing an innovative algorithm to map the emotional state and passenger safety risk to generate practical and effective alarms.
• Incorporating Machine Vision at the Edge (Mez) technology to overcome the challenges caused by network latency in real-time video transmission.
The rest of the paper has been organized into six different sections. The second section discusses the recent relevant research. The methodology has been presented in the third section. The fourth section highlights the experimental results and evaluation. The limitation and future scope of this experiment have been presented in the fifth section. Finally, the paper has been concluded in the sixth section.

Literature review
The Internet of Things, abbreviated as IoT, is a physical device consisting of sensors, processors, software, and other objects. The core responsibility of IoT devices is to connect the device to the internet to exchange data over the internet [11]. Although the concept of IoT is simple, the application domain is enormous [12]. It is applied in smart homes [13], autonomous vehicles [14], the healthcare sector [15], industrial control [16], agriculture [17], etc. Autonomous vehicles and transport management are the two most widely explored areas of IoT applications [18], [19]. The research presented in this paper has taken a different approach to IoT applications in the transport industry. The innovative methodology of this paper is about passenger safety using IoT devices. T. Boshita et al. [20] demonstrated a creative application of IoT in bus location systems using LoRaWAN. The core focus of this paper is to improve the quality of the bus network for efficient management. They used 3G/LTE network to maintain communication. The proposed IoT device of this paper uses a similar communication technology. However, it focuses on passenger safety. S. Geetha et al. showed an IoT-based intelligent bus transport system. Their research approach is passenger-centered. This system tracks the information related to bus numbers, arrival time, and onboard passenger count and generates intelligent reports for the users. The prototype model developed in their research demonstrates acceptable performance. However, passenger safety is not within the context of this study [21], which has been explored in the proposed methodology. Combining Deep Learning (DL) algorithms with IoT is a popular field of research [22]. J. Jabamony et al. combine the Artificial Neural Network (ANN) with IoT devices to intelligently predict the bus arrival time [23]. The methodology of the proposed paper uses the IoT-CNN combination, which is similar to [23]. However, the core focus of it is to strengthen passenger safety. Most IoT applications in bus transportation are in location tracking, arrival time prediction, service quality improvement, and management [24]–[27]. To our best knowledge, the proposed paper is the first application of IoT in passenger safety. The proposed paper combines signals from IR sensors with facial emotion recognition technology to strengthen passenger safety in a school bus. Facial emotion recognition using Convoltuional Neural Network (CNN) is a well-developed field of research [28], [29]. I. Adjabi et al. presented the evolution and current maturity level of this technology [30]. A survey by [31] by F. Z. Canal et al. shows the limited scope of contribution in this domain. This paper explores the potential of CNN to recognize facial emotions from two different contexts and overcomes its associated challenges. The first context is the using CNN in an IoT environment where the computational resource is limited. And the second context is developing the relationship between emotional state and passenger safety.

R. Pathak et al. developed a system to recognize a baby's facial expression using deep learning and IoT edge computing. Their system recognizes three motions - (i) Happy, (ii) Crying, and (iii) Sleeping [32]. We used a similar approach. However, the proposed system is developed based on 26 different emotional states. An innovative approach by M. S. Hossain et al. uses edge cloud and CNN to recognize emotion from image and speech features. Their system recognizes emotions with 84.95% accuracy on an average [33]. A social IoT framework based facial expression recognition system has been developed by S. Barra et al. [34], which is applied in the medical sector. None of these methods explicitly addressed the solution to latency issues in real-time facial emotion recognition using IoT devices. The proposed methodology presents a novel application of Mez technology to overcome this challenge [35].

Methodology
A. The IoT Device
The IoT device developed and presented in this paper, illustrated in figure 1, consists of four elements. Among these elements, two elements are sensors, one is a communication device, and another is a single-board computer. The sensors are the Infrared (IR) sensor and Pi Camera. The communication device used in this paper is the SIM7600 GSM GPS 4G LTE module. And the

single-board computer is a Raspberry Pi. The system architecture of the implementation model of the proposed IoT-based bus security system has been prepared to implement according to the design presented in figure 2. The symbol has been made larger than the appropriate size on the figure to highlight its position. It becomes easier to understand the overall architecture of the proposed system because of the large noticeable symbols of the sensors, communication devices, and single-board computers. There are two IR emitters at two sides of the bus mounted at the end of the bus. These emitters are aimed at the IR receivers installed at the front left and right sides. The IR receiver is connected to the Raspberry pi. The pi camera is mounted at the front of the bus with an angle of 60 degrees with respect to the vertical support of the driver's seat.

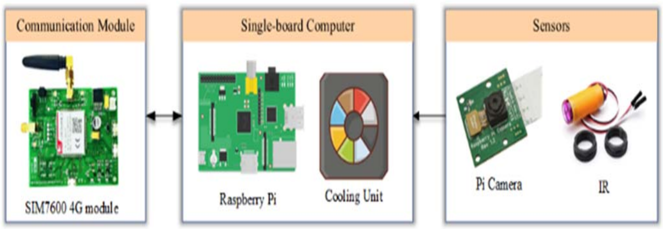

Fig.1. The overview of the IoT device developed and used for this research

1) IR Sensor: : A single-beam Infrared (IR) sensor, named ABO-20, has been used in this experiment. It has been illustrated in figure 3. Using IR sensors on the road is challenging. There are random dynamic and static light sources with multiple intensities and colors on the road. These lights interfere with the IR receiver. As a result, there is a probability of having an incorrect response from the sensor. One of the reasons behind using the ABO-20 is its 50,000 LUX anti-glare feature. The ABO-20 IR sensor runs at 12V DC with a 15 mA current at the transmitter and a 30 mA current at the receiver. This sensor has both Normally Open Circuit (NOC) and Normally Closed Circuit (NCC) switching modes. In this experiment, the NOC switching mode has been used to save energy. It generates a signal only when IR transmission is interrupted. It has an effective

range of 2 to 15 meters. However, it detects designated infrared signals up to 20-meter distance outdoors. The operating IR frequency generated by the transmitter is 1.92Khz with a wavelength of 9.40×10−11 meter. The sensor used in this experiment does not have temperature sensitivity within the range of −30oC to 50oC.

2)

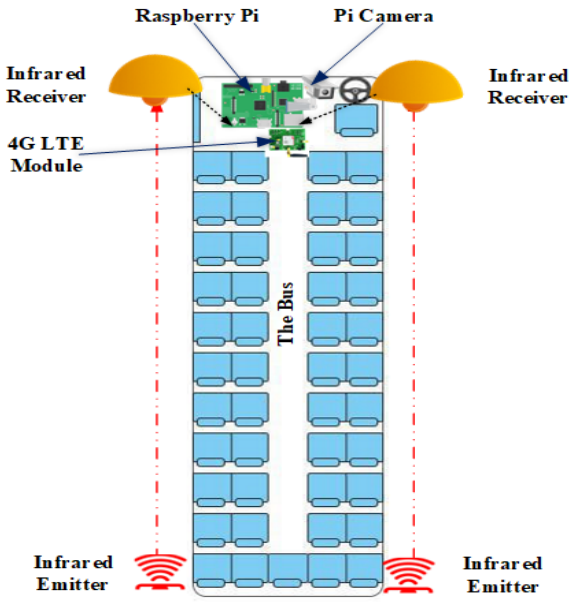

Fig.. 2. System architecture

2) Communication Device: The SIM7600 GSM GPS 4G LTE module, which is illustrated in figure 4, has been used in this research to connect the microcontroller to the internet. It has uplinks up to 50 Mbps and downlinks up to 150 Mbps. It supports multiple communication bands. This experiment used Quad-Band TDD-LTE B38/B39/B40/B41 to maintain the communication of the microcontroller with the internet. We used a 5.0V power supply through a USB cable to run it. The AT Commands have been used to control the GSM module.

Fig. 3.The ABO-20 single beam IR sensor

3) Raspberry Pi: The Raspberry Pi 2 Model B, illustrated in figure 4, has been used in this experiment. It is powered by a 900MHz quad-core ARM Cortex-A7 CPU. The primary memory of the experimenting device is 1GB. A 32GB Micro Secured Digital (MicroSD) memory card has been used in the device. It has up to 100MB/s reading and writing speed. The 32-bit Raspberry Pi Operating System (OS) is installed in this memory card. The excessive heat issue of the Raspberry Pi was discovered during the experimental phase. The GeekPi cooling fan has been used along with an additional heatsink to mitigate the overheating effect.

4) Pi Camera: The Pi camera v1.3, illustrated in figure 5, has been used in this experiment as the imaging device. It has a native 5 Mega Pixel (MP) sensor size. It supports recording at 1080P, 720P, and 960P at 30 Frame Per Second (FPS), 60 FPS, and 45 FPS, respectively. It comes with a maximum 6-second exposure time. The focal ratio of the experimenting camera is F2.9. The approximate Vertical Field of View (FoV) is 41.41 ± 0.11 degree. It has an approximate 53.50 ± 0.13 degree horizontal Field of View (FoV). It is a fixed-focus camera with a pixel size 1.4µm × 1.4µm. The Camera Serial Interface has been used in this research to connect the camera to the Raspberry Pi through a 15-pin ribbon cable.

Fig. **4**. The SIM7600 GSM GPS 4G LTE module with the Raspberry Pi

Fig. 5. The Pi Camera connected to the Raspberry Pi

A. Driver's Attention Detection Module

1) Dataset: The EMOTIC dataset has been used in this experiment to train a Convolutional Neural Network (CNN) to recognize the facial expression of the driver. The facial emotions available in the dataset have been prepared from an uncontrolled environment. That is why it has images of natural expressions of human faces. The EMOTIC dataset consists of 23,571 images. There are 34,320 annotations on these images. The images are classified into 26 discrete categories. These 26 categories have been grouped into three classes based on the level of danger during driving. These three classes are - (i) No, (ii) Medium, and (iii) High. These classes have been encoded with numeric values.

The dataset description, classes, and their codes have been listed in table I.

Table 1. Dataset description and class encoding

Serial	Class	Dangerous	Code
1	Affection	No	0
2	Anger	High	2
3	Annoyance	High	2
4	Anticipation	No	0
5	Aversion	High	2
6	Confidence	No	0
7	Disapproval	No	0
8	Disconnection	High	2
9	Disquietment	High	2
10	Doubt/Confusion	Medium	1
11	Embarrassment	No	0
12	Engagement	No	0
13	Esteem	No	0
14	Excitement	No	0
15	Fatigue	High	2
16	Fear	No	0
17	Happiness	No	0
18	Pain	High	2
19	Peace	No	0
20	Pleasure	No	0
21	Sadness	Medium	1
22	Sensitivity	Medium	1
23	Suffering	Medium	1
24	Surprise	No	0
25	Sympathy	No	0
26	Yearning	No	0

The images of the EMOTIC dataset are not ready for the network developed in this research project. That is why we have preprocessed the images according to the requirements of the Convolutional Neural Network (CNN) in this paper. First, the images have been resized to 224 × 224 × 3 using equation 1. The ratio of the images is preserved while resizing. As a result, the maximum amount of features is preserved.

(1) $$(m', m') = \frac{M}{max(m, n)} \times (m, n)$$

Image features are essential elements in training CNNs. Poor feature quality hampers the performance of CNN-based classifiers. It is evident that without an adequate amount of rich features, a CNN cannot reach to its maximum potential with large datasets. That is why we have enhanced the image features using a histogram equalizer defined by the equation 2.

(2) $$h(v) = round(\frac{cdf(v) - cdf_{min}}{(M \times N) - cdf_{min}} \times (L - 1))$$

The equation 2 requires the Cumulative Distribution Function (CDF). The cdf of equation 2 is defined by equation 3 where M is the width of the images. And the N represents the height of the images.

(3) $$cdf(i) = \sum_{j=0}^{i} p(x = j)$$

After processing the dataset, they are stored in directories named after respected classes. There are 26 classes in the EMOTIC dataset. That is why there are 26 directories in the experimental environment. The algorithm 1 has been used in this experiment to resize images, process them after resizing, and finally store them in appropriate directories

Algorithm 1 The EMOTIC Dataset Preparation
Require: D_t : *Dataset*;S_t : *LocalDirectory*;
Variables: M : *BufferMemory*;L : *ImageLabel*;
$I_n \leftarrow length(D_t)$ ▷ Counting images
for $(i = 0; i \leq I_n; i++)$ do
$[M,L] \leftarrow D_t[i]$
$M \leftarrow resize(M,[227,227],RGB)$ ▷ Image Resizing
$M \leftarrow hist(M)$ ▷ Histogram processing
$Dir \leftarrow list(S_t)$ ▷ Getting directory list
if $Dir == L$ then
    $S_t \leftarrow write(M,L)$ ▷ Writing images
else
    $create(L)$ ▷ Creating directories
$S_t \leftarrow write(M,L)$ ▷ Writing images in directories
end if
end for

2) Network Architecture: We designed and implemented a Convolutional Neural Network (CNN) to learn from the processed dataset presented in section III-B1. It has been illustrated in figure 6. The proposed network is a 50-layerdeep CNN. It has been designed to classify facial expressions from the EMOTIC dataset. The mathematical expression of the proposed CNN is defined in equation 4 where σ represents the activation function, Ii means the input signal, the w stands for the weight vector, and the b represents the bias.

$$\sigma(I_i) = \sigma(w * I_i + b)$$

It has been observed during the experiment that the vanishing gradient problem impacts the feature learners. We used the concept of a bypass layer to solve this problem. The bypass layer is defined by equation 5. It enables the input signal to bypass designated layers. As a result, the existing layer learns from the image, and the subsequent layer receives the uncompressed image. That is why the deep layers do not suffer from vanishing gradient problems. The equation 5 is used to bypass the $(l - i)^{th}$ layer's signal to l^{th} layer.

(5) $$\sigma(I_i) = \sigma(I_i) + I_i$$

a) Batch Normalization: The EMOTIC dataset has 23,571 images. Attempt to train the proposed CNN with all of these image all at once is impractical. Grouping training images into small batches and train the CNN using these batches is more efficient and practical approach. We used the batch normalization method with a mini-batch size of 64 in in this research. The batch normalization method is defined by equation 6.

(6) $$Y_i = \frac{(x_i - \mu)}{\sqrt{\sigma^2 + \epsilon}}\theta + b$$

In the mathematical expression of the batch normalization method, the μ is the mean, and σ represents the variance which are expressed by equation 7 and 8, respectively. The minibatch size we used in this experiment is 64. That is why the maximum limits in equations 7 and 8 are 64.

(7) $$\mu = \frac{1}{64}(\sum_{i=1}^{64} x_i)$$

(8) $$\sigma = \frac{1}{64}(\sum_{i=1}^{64} x_i - \mu)^2$$

(8) *b) Activation*

b) Activation Function: The signals from every node of the CNN is mapped in between the probability range. The lowest and highest probability is 0 and 1, respectively. It is essential to map the signals of the nodes of CNN layers within this range. The Rectified Linear Unit (ReLU) defined by equation 9, is an ideal mathematical model to map input signal within the range between 0 to 1. That is why it has been used as the activation function of the hidden nodes of the proposed network.

(9) $$Yi,j,k = max\{0,xi,j,k\}$$

(10) $$\mathrm{Lt} = \psi \mathrm{y\ i} + \log \sum_j exp(\psi j)$$

In equation 9, *i* and*j* represent the width and height of the input image. That means the input to this equation is a 2D matrix. This matrix is expressed by *k*.

c) Pooling Layer: The network architecture illustrated in figure 6 uses max-pooling after the convolution layer to reduce the feature vector size. It is evident that a large number of relevant features enhances the performance of the CNNs. However, it comes with the cost of computational resources. If the feature vector is too large, it takes much longer to train the network. At the same time, there are risks of overfitting the network. That is why max-pooling, defined by equation 10, has been used in this experiment.

(11) $$Y^i = \max(0, \sum_i^p x^{i-1}(2 \times 2)^i)$$

It accelerates the learning process and suppresses and reduces the probability of overfitting. The Y^i in equation 10 is the output of the max-pooling layer. Here,x^{i-1} is the input feature vector. The size of the pooling layer is expressed by *p*. The convolution operation, which is done through batch normalization and processed by ReLU, is the input to the max-pooling layer. The equation 11 expresses the overall maxpooling process.

(12) $Y_{Pool} = MaxPooling(ReLU(B_{norm}(Conv(x,\omega))))$

d) Softmax Classifier and Loss Function: The final classification layer of the proposed CNN uses a log-loss-based softmax function. There are 26 classes in the dataset. This layer maps the input image to any of these classes with a confidence score. The confidence score is the probability of the correct prediction expressed in percentage. The operation of the softmax layer is defined by equation 12.

(13) $$P_l = \frac{exp^{\psi_{y_i}}}{\sum_j exp(\psi_j)}$$

The equation 13 finds out the total loss (L_t). The ψ represents the class, and the ψ_j stands for the elements of the class. The goal of the proposed CNN is to reduce the difference between the ground truth and the prediction. It is done by minimizing the loss calculated by equation 13.

Learning Algorithm: The proposed CNN has been trained using the back-propagation algorithm.The backpropagation algorithm requires the cost function, which has been defined by equation 14. The learning algorithm iteratively reduces the difference between the predicted class and the ground truth.

(14) $$C_f = -\frac{1}{|x|}\sum_{i=1}^{|x|} ln(P(y^i|x^i|))$$

Updating the weights of the hidden nodes of the CNN is not an optimal way to train the CNN. We used the Adaptive Moment Estimation (ADAM) optimizer to update weights effectively. It uses the cost function defined by equation 14. The ADAM optimizer is defined by equation 15.

(15) $$m_t = \beta_1 m_{t-1} + (1-\beta_1)[\frac{\delta C_f}{\delta \omega_t}]$$

The equation 15 finds out the momentum, which is later used to update the weight. Here, m_t is the momentum, which is then aggregated at time *t*. The β is a constant parameter that has been set to 0.9 in this experiment. The momentum calculation requires the derivation of the cost function, which is expressed as $\frac{\delta C_f}{\delta \omega_t}$. The m_t measured from 15 is used in equation 16 to calculate the optimal weight to update.

(16) $\omega_{t+1} = \omega_t - \alpha m_t$

In equation 16, the α represents the learning rate. The learning rate is used to control the modification of the weight. It ranges from 0 to 1. The ω_t is the weight to be updated. And ω_{t+1} is the weight after updating.

C. Latency Control through Mez

1) Mez System Architecture: The original Mez system architecture developed by A. George et al. has been designed to connect multiple IoT camera nodes simultaneously. The Edge server is connected to it through a wireless network [7]. The proposed system has only one camera. The Edge server and the camera are connected through a wired network. Unlike the original Mez system, the proposed system uses a licensed 4G band to communicate with the cloud server. That is why a modified Mez architecture, illustrated in figure 7, has been

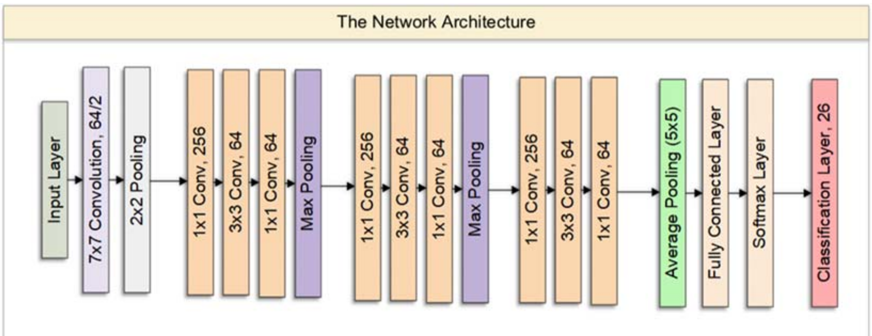

FIG 6. The Convolutional Neural Network used in this research

Table 2. ASvailable knob settings, their roles, effect and scope

Knob	Role	Frame Size Reduction	Scope
1	Resolution Adjustment	84%	Resolutions: 1312x736, 960x528, 640x352, and 480x256
2	Colorspace Modification	62%	Colorspaces: BGR, Grayscale, HSV, LAB, and LUV
3	Blurring	46%	Kernel size: 5x5, 8x8, 10x10, and 15x15
4	Artifact Removal	98%	Contour-based approach
5	Frame Difference	40%	Linear frame difference based method

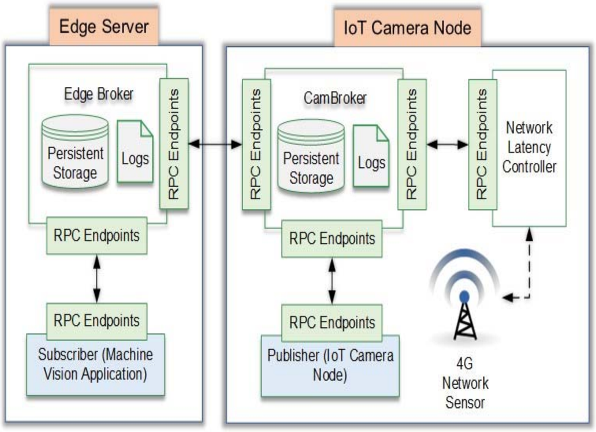

Fig. 7. The modified Mez architecture

used in this experiment. This architecture has a 4G network sensor to monitor the quality of the network. It communicates with the Network Latency Controller.

2) Latency VS. Quality Trade-off: The proposed system utilizes the latency VS. quality trade-off capability of Mez technology. There are five different knob settings to adjust the frame quality according to the application accuracy requirements. The available knob settings, their roles, the effect on frame size reduction, and the application scopes have been listed in table II.

Experimental results and evaluation

A. Sensor Response Analysis

The proposed IoT-based passenger safety system's performance depends on appropriate sensor response. The ABO20 IR receiver sensor used in this experiment tolerates up to 50,000 Lux external interference, according to the manufacturer's documentation. However, this documentation does not guarantee that the experimenting unit will demonstrate the same response. That is why we experimented with the proposed system in two different setups. The first experimental setup is in daylight, and the second setup is at night. The responses from the sensor were observed for 30 minutes in both experimental environments. Along with luminance variations, we studied the sensor response as well. The evaluation of the experiment has been illustrated in figure 8. The sensor response of figure 8 shows that external interference does not impact the sensor response. We operated the sensor at NOC switching mode. No current should flow during the OFF state of the sensor, and maximum current should flow when the sensor sends a signal. It has been observed that a marginal leakage current of the microampere range flows through the circuit all the time. However, the ON state of the sensor generates current at the milli-ampere range, which is 100 times more than the leakage current. That is why the system response is effortlessly identifiable. Based on the sensor response analysis, we conclude that the experimenting system generates appropriate responses.

B. IoT Response Delay

IoT devices related to passenger safety must communicate in real time. However, there are always response delays because of multiple factors. The proposed IoT device is installed in a bus that moves at varying speeds and comes under the coverage of different Base Transceiver Stations (BTS). It causes random delays. A subset of the IoT response relay analysis data has been listed in table III. It shows the response timestamp of events, the timestamp of corresponding reflection in the cloud server, the timestamp of receiving the acknowledgment bit, and the overall delay in milliseconds. It is noticeable that there is no regular pattern in system response delay from figure 9. The response time is the timestamp where the event occurred. The reflection to the cloud server exhibits linear characteristics. There is a similar pattern between acknowledgment and delay. However, it does not maintain any regular, predictable structure. The average delay calculated from 340 random events over 2 hours is 113 milliseconds, considered an acceptable real-time response.

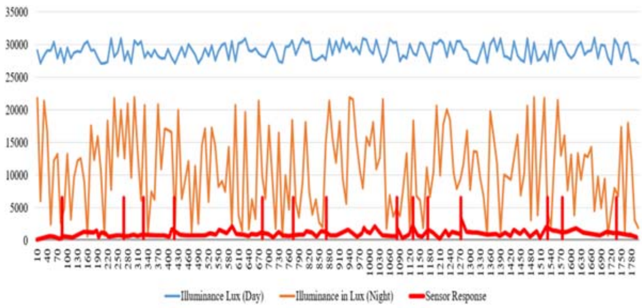

Fig. 8.The evaluation of the sensor response in different environments

Table 3. Response delay analysis of the IOT device

Event	Response (ms)	Reflection (ms)	Acknowledgement (ms)	Delay (ms)
1	300	381	337	29
2	735	738	764	32
3	928	931	960	218
4	1128	1133	1346	28
5	1842	1916	1870	426
6	2430	2433	2856	21
7	2674	2681	2695	223
8	2849	2858	3072	239
9	3053	3056	3292	52
10	3604	3608	3656	

C. Attention Detection Performance

The learning curve illustrated in figure 10 shows the learning process of the experimenting CNN. Figure 10(a) shows progress's training and validation accuracy. and figure 10(b) presents the training and validation loss. The training process takes 62 minutes and 14 seconds. With 500 iterations, the network learns to predict facial emotion with 84.95% training and 83.73% validation accuracy. The overall validation loss of the network during the training phase is 11.48%, and the training loss is 14.55%. The performance of the trained network has been evaluated using the testing dataset and real-world scenarios. We used state-of-the-art machine learning evaluation metrics. The literature review shows that accuracy, precision, recall, and F1-score are the most widely used evaluation metrics. The mathematical expression of these metrics has been presented by equation 17, 18, and 19, respectively [35]. These values are calculated using the True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) from the confusion matrix illustrated in figure 11.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

We used K-fold cross-validation at $k = 5$ while evaluating the performance of the proposed system with the testing dataset. The value of the evaluation metrics at different k has been listed in table IV. The proposed network's average accuracy, precision, recall, and error rate on the testing dataset are 83.36%, 86.65%, 85.21%, and 16.64%.

Table 4. Performance on emotion analysis

K	Accuracy	Precision	Recall	Error Rate
1	82.19	86.95	85.17	17.81
2	84.57	85.95	85.14	15.43
3	83.58	86.76	85.42	16.42
4	83.99	86.92	85.04	16.01
5	82.47	86.65	85.3	17.53
Average	83.36	86.65	85.21	16.64

D. Latency Reduction

The experiment to study latency sensitivity and reduction rate has been conducted in two phases. Each phase is 60 minutes long. No specific bus route has been selected to maintain the experiment's integrity. The real-world scenario is always random. Experimenting on a random route reflects the real-world randomness in the experimental setup. The first experimental phase records the latency in 10 seconds intervals without the Mez technology. The second phase records at the same interval as Mez. The findings have been illustrated in figure 12. The average latency of the proposed system without the Mez is 8.03 seconds. After applying the Mez, it reduces to 1.35 seconds. That means Mez reduces the latency by 83.19%.

Limitation and future scope

The proposed IoT-based real-time passenger safety system works perfectly in laboratory and practical settings. However, no system is immune to limitations. The proposed system is not an exception. The limitations of this experiment have been considered the opportunity to conduct further research and improve the system's quality.

A. Effect of Rain

The sensors are installed outside of the school bus. That means it is exposed to the natural environment. The experimental analysis and performance evaluation were conducted in the dry season. Creating artificial rain was out of the scope of resource limitation. One of the paper's weaknesses is that the effect of rain on the system is still unknown. However, this limitation opens the door to further experiments about the seasonal effect on the proposed system.

B. Network Switching

The SIM7600 GSM GPS 4G LTE module has been used in this for communication. It supports 2G, 3G, and 4G networks. However, the proposed system has been designed to operate on a 4G network only for convenience. The system's response in the 2G and 3G networks remains unknown. Network switching is an effective way to maintain seamless connection in wireless mobile communication. Depending on the availability, the mobile devices switch among the 2G, 3G, and 4G networks. The proposed system does not have such flexibility, which is a major limitation of the paper. Introducing such flexibility is more complicated than it seems, which paves to path to conducting more experiments to study the effect of Mez on different types of networks.

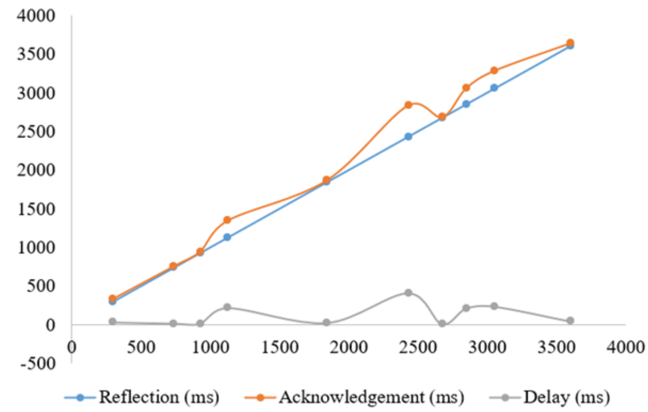

Fig. 9. The delay analysis of the proposed IoT device

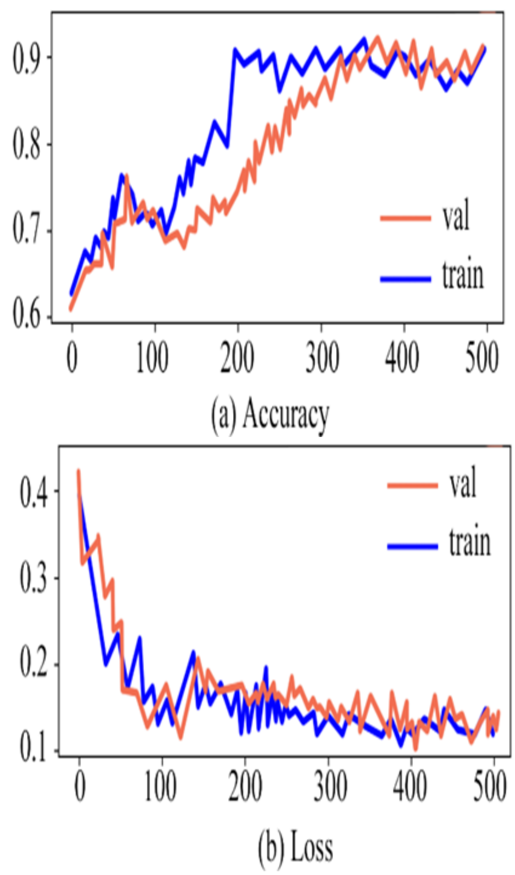

Fig. 10.The learning curve

C. Video Stabilization The bus shakes depending on the condition of the road. The video becomes shaky, and the system's accuracy is compromised when it happens. There are different video stabilization algorithms. However, resource constraints make it impossible to use them in the presented IoT device. Developing an IoT device-compatible video stabilization algorithm is a potential solution to this problem which creates the opportunity to conduct more experiments.

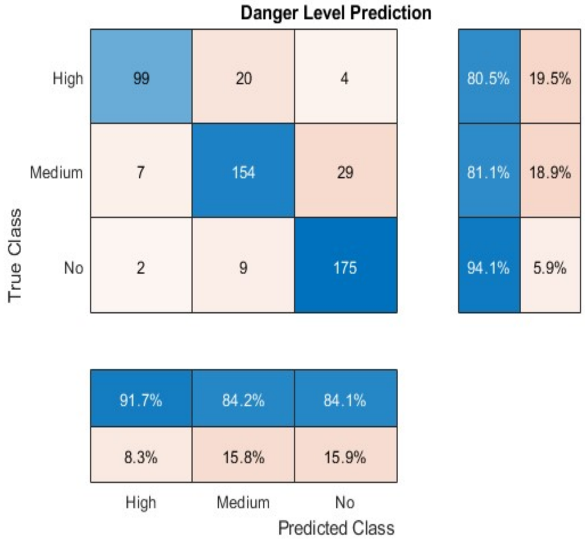

Fig. **11**.The confusion matrix

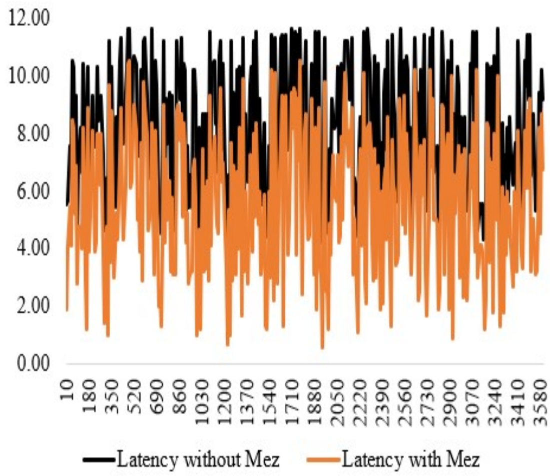

Fig. **12**.The latency reduction after using the Mez

Conclusion

The widespread applications of IoT devices have made our lives better and easier. Despite the immense potential, the application domain of IoT in passenger safety in school buses is not well-developed. It is a common scenario in school buses that the children open the window and put their hands, heads, or upper part of the body outside to enjoy the wind flow. It is a dangerous move, and accidents can happen at any time. Solely depending on the driver to drive and maintain discipline simultaneously distracts the driver, which imposes a bigger safety concern. Moreover, a driver may not always be in an emotionally stable condition to properly maintain the passenger's safety by driving safely. We have presented an innovative IoT-based passenger safety system to strengthen safety in school buses by generating alarms and notifying the concern to the respected authority over the internet. At the same time, this paper combines the CNN-based emotion recognition module to scan the emotional state of the driver and predict the potential risks. The resource constraint in IoT imposes additional challenges in real-time video processingbased applications. This paper has presented the application of Mez to maintain application frame quality VS. latency tradeoff to maintain acceptable system accuracy. According to the experimental analysis and numerical performance indicator, the proposed system is an effective, robust, and practical passenger safety system.

***Authors**: **Ali Abdulaziz Alawad** He is a master student in communication engineering in university of Hail, Saudi Arabia. He got his bachelor in university of Hail in 2018. He is interest in wireless locating system.*
Email: eng.aliabdulaziz@gmail.com
***Muhammad Tajammal Chughtai** Department of Electrical Engineering, University of Hail, Saudi Arabia Email: mt.chughtai@uoh.edu.sa*

REFERENCES

[1] Z. S. Khan, W. He, and M. Menendez, "Application of modular vehicle´ technology to mitigate bus bunching," *Transportation Research Part C: Emerging Technologies*, vol. 146, p. 103953, 2023.

[2] C. Utary, D. S. Nabahan, E. Budianto, and S. N. I. Sari, "Planning of school safety zone (zoss) on the education road of merauke regency," in *MATEC Web of Conferences*, vol. 372, p. 07004, EDP Sciences, 2022.

[3] J. Hinton, B. Watson, and O. Oviedo-Trespalacios, "A novel conceptual framework investigating the relationship between roadside advertising and road safety: the driver behaviour and roadside advertising conceptual framework," *Transportation research part F: traffic psychology and behaviour*, vol. 85, pp. 221–235, 2022.

[4] J. Ahmed, N. Ward, J. Otto, and A. McMahill, "How does emotional intelligence predict driving behaviors among non-commercial drivers?," *Transportation research part F: traffic psychology and behaviour*, vol. 85, pp. 38–46, 2022.

[5] S. Saurav, P. Gidde, R. Saini, and S. Singh, "Dual integrated convolutional neural network for real-time facial expression recognition in the wild," *The Visual Computer*, pp. 1–14, 2022.

[6] Z. Lei, S. Ren, Y. Hu, W. Zhang, and S. Chen, "Latency-aware collaborative perception," in *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXII*, pp. 316–332, Springer, 2022.

[7] A. George, A. Ravindran, M. Mendieta, and H. Tabkhi, "Mez: An adaptive messaging system for latency-sensitive multi-camera machine vision at the iot edge," *IEEE Access*, vol. 9, pp. 21457–21473, 2021.

[8] S. Huai, D. Liu, H. Kong, W. Liu, R. Subramaniam, C. Makaya, and Q. Lin, "Latency-constrained dnn architecture learning for edge systems using zerorized batch normalization," *Future Generation Computer Systems*, vol. 142, pp. 314–327, 2023.

[9] Y.-C. Chen, E. M. Nahum, R. J. Gibbens, D. Towsley, and Y.-s. Lim, "Characterizing 4g and 3g networks: Supporting mobility with multipath tcp," *School of Computer Science, University of Massachusetts Amherst, Tech. Rep*, vol. 22, 2012.

[10] A. Imteaj, U. Thakker, S. Wang, J. Li, and M. H. Amini, "A survey on federated learning for resource-constrained iot devices," *IEEE Internet of Things Journal*, vol. 9, no. 1, pp. 1–24, 2021.

[11] D.-R. Berte, "Defining the iot," in *Proceedings of the international conference on business excellence*, vol. 12, pp. 118–128, 2018.

[12] A. Colakoviˇ c and M. Had´ zialiˇ c, "Internet of things (iot): A review of´ enabling technologies, challenges, and open research issues," *Computer networks*, vol. 144, pp. 17–39, 2018.

[13] L. Babangida, T. Perumal, N. Mustapha, and R. Yaakob, "Internet of things (iot) based activity recognition strategies in smart homes: A review," *IEEE Sensors Journal*, 2022.

[14] S. Jha, N. Jha, D. Prashar, S. Ahmad, B. Alouffi, and A. Alharbi, "Integrated iot-based secure and efficient key management framework using hashgraphs for autonomous

vehicles to ensure road safety," *Sensors*, vol. 22, no. 7, p. 2529, 2022.
[15] N. Faruqui, M. A. Yousuf, M. Whaiduzzaman, A. Azad, A. Barros, and M. A. Moni, "Lungnet: A hybrid deep-cnn model for lung cancer diagnosis using ct and wearable sensor-based medical iot data," *Computers in Biology and Medicine*, vol. 139, p. 104961, 2021.
[16] H. T. Truong, B. P. Ta, Q. A. Le, D. M. Nguyen, C. T. Le, H. X. Nguyen, H. T. Do, H. T. Nguyen, and K. P. Tran, "Light-weight federated learning-based anomaly detection for time-series data in industrial control systems," *Computers in Industry*, vol. 140, p. 103692, 2022.
[17] M. S. Farooq, S. Riaz, A. Abid, T. Umer, and Y. B. Zikria, "Role of iot technology in agriculture: A systematic literature review," *Electronics*, vol. 9, no. 2, p. 319, 2020.
[18] S. Muthuramalingam, A. Bharathi, S. Rakesh Kumar, N. Gayathri, R. Sathiyaraj, and B. Balamurugan, "Iot based intelligent transportation system (iot-its) for global perspective: A case study," *Internet of Things and Big Data Analytics for Smart Generation*, pp. 279–300, 2019.
[19] M. F. Elahe, M. Jin, and P. Zeng, "Knowledge-based systematic feature extraction for identifying households with plug-in electric vehicles," *IEEE Transactions on Smart Grid*, vol. 13, no. 3, pp. 2259–2268, 2022.
[20] T. Boshita, H. Suzuki, and Y. Matsumoto, "Iot-based bus location system using lorawan," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pp. 933–938, IEEE, 2018.
[21] S. Geetha and D. Cicilia, "Iot enabled intelligent bus transportation system," in *2017 2nd International Conference on Communication and Electronics Systems (ICCES)*, pp. 7–11, IEEE, 2017.
[22] A. Awajan, "A novel deep learning-based intrusion detection system for iot networks," *Computers*, vol. 12, no. 2, p. 34, 2023.
[23] J. Jabamony and G. R. Shanmugavel, "Iot based bus arrival time prediction using artificial neural network (ann) for smart public transport system (spts)," *International Journal of Intelligent Engineering & Systems*, vol. 13, 2019.
[24] K. Sridevi, A. Jeevitha, K. Kavitha, K. Sathya, and K. Narmadha, "Smart bus tracking and management system using iot," *Asian Journal of Applied Science and Technology (AJAST) Volume*, vol. 1, 2017.
[25] Y. Zhu, F. Yan, J.-S. Pan, L. Yu, Y. Bai, W. Wang, C. He, and Z. Shi, "Mutigroup-based phasmatodea population evolution algorithm with mutistrategy for iot electric bus scheduling," *Wireless Communications and Mobile Computing*, vol. 2022, 2022.
[26] R. S. Krishnan, A. Kannan, G. Manikandan, S. S. KB, V. K. Sankar, and K. L. Narayanan, "Secured college bus management system using iot for covid-19 pandemic situation," in *2021 third international conference on intelligent communication technologies and virtual mobile networks (ICICV)*, pp. 376–382, IEEE, 2021.
[27] A. Rahmatulloh, F. M. Nursuwars, I. Darmawan, and G. Febrizki, "Applied internet of things (iot): the prototype bus passenger monitoring system using pir sensor," in *2020 8th International Conference on Information and Communication Technology (ICoICT)*, pp. 1–6, IEEE, 2020.
[28] W. Mellouk and W. Handouzi, "Facial emotion recognition using deep learning: review and insights," *Procedia Computer Science*, vol. 175, pp. 689–694, 2020.
[29] M. S. Arman, M. R. Alam, H. Jahan, L. Islam, M. H. Sammak, and K. B. M. B. Biplob, "A data mining approach to finding face mask from bangladeshi news channel," in *2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pp. 1–7, IEEE, 2022.
[30] I. Adjabi, A. Ouahabi, A. Benzaoui, and A. Taleb-Ahmed, "Past, present, and future of face recognition: A review," *Electronics*, vol. 9, no. 8, p. 1188, 2020.
[31] F. Z. Canal, T. R. Muller, J. C. Matias, G. G. Scotton, A. R. de Sa Junior,¨ E. Pozzebon, and A. C. Sobieranski, "A survey on facial emotion recognition techniques: A state-of-the-art literature review," *Information Sciences*, vol. 582, pp. 593–617, 2022.
[32] R. Pathak and Y. Singh, "Real time baby facial expression recognition using deep learning and iot edge computing," in *2020 5th International conference on computing, communication and security (ICCCS)*, pp. 1– 6, IEEE, 2020.
[33] M. S. Hossain and G. Muhammad, "Emotion recognition using secure edge and cloud computing," *Information Sciences*, vol. 504, pp. 589– 601, 2019.
[34] S. Barra, S. Hossain, C. Pero, and S. Umer, "A facial expression recognition approach for social iot frameworks," *Big Data Research*, vol. 30, p. 100353, 2022.
[35] S. Trivedi, N. Patel, and N. Faruqui, "Ndnn based u-net: An innovative 3d brain tumor segmentation method," in *2022 IEEE 13th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, pp. 0538–0546, IEEE, 2022.