Human Activity Recognition-based Cyber-Physical System Security Through IoT-Cloud Computing

Sandesh Achar *, Nuruzzaman Faruqui *, Md Whaiduzzaman *, Albara Awajan, Moutaz Alazab

Posted Date: 10 March 2023

doi: 10.20944/preprints202303.0183.v1

Keywords: Cyber-physical security; Human activity recognition; GoogleNet; BiLSTM; Deep Learning; Algorithm

Article

Human Activity Recognition-Based Cyber-Physical System Security through IoT-Cloud Computing

Sandesh Achar [1,*,†], Nuruzzaman Faruqui [2,*], Md Whaiduzzaman [3,*], Albara Awajan [4] and Moutaz Alazab [5]

[1] sandeshachar26@gmail.com
[2] faruqui.swe@diu.edu.bd
[3] md.whaiduzzaman@qut.edu.au
[4] m.alazab@bau.edu.Jo
[5] a.awajan@bau.edu.jo
* Correspondence: sandeshachar26@gmail.com (S.A.); faruqui.swe@diu.edu.bd (N.F.); md.whaiduzzaman@qut.edu.au (M.W.)
† Senior Manager- Software Engineering, Walmart Global Tech, Sunnyvale, California 94086 (email: sandeshachar26@gmail.com)

Abstract: Cyber-physical security is vital for protecting key computing infrastructure against cyber attacks. Individuals, corporations, and society can all suffer considerable digital asset losses due to cyber attacks, including Data loss, theft, financial loss, reputation harm, company interruption, infrastructure damage, ransomware attacks, and espionage. A cyber-physical attack harms both digital and physical assets. Cyber-physical system security is more challenging than software-level cyber security because it requires physical inspection and monitoring. This paper proposes an innovative and effective algorithm to strengthen Cyber-Physical Security (CPS) with minimal human intervention. It is a Human Activity Recognition (HAR)-based approach where a GoogleNet-BiLSTM network hybridization has been used to recognize suspicious activities in cyber-physical infrastructure perimeter. The proposed HAR-CPS algorithm classifies suspicious activities from real-time video surveillance with an average accuracy of 73.15%. It incorporates Machine Vision at the IoT Edge (Mez) technology to make the system latency tolerant. Dual-layer security has been ensured by operating the proposed algorithm and GoogleNet-BiLSTM hybrid network from a cloud server, which ensures the security of the proposed security system. The innovative optimization scheme makes it possible to strengthen cyber-physical security with $4.29 per month only.

Keywords: cyber-physical security; human activity recognition; GoogleNet; BiLSTM; deep learning; algorithm

1. Introduction

The field of cyber security that deals with the security of physical computing devices is called cyber-physical security. A wide range of devices, for example, desktops, laptops, servers, network switches, routers, Internet of Things (IoT), fall under the category of cyber-physical systems. As a matter of fact, every physical system associated with computing is a subset of cyber-physical systems [1]. Cyber-physical security is critical because attacks on these systems can have serious consequences, including hardware damage, service interruption, malware injection through physical ports, and data disclosure. Cybersecurity is incomplete without cyber-physical security. Organizations take various measures to protect both digital and physical assets. However, guarding physical assets for 24×7 is much more challenging than digital assets [2]. The Human Activity Recognition-based Cyber-Physical Security (HAR-CPS) algorithm presented in this paper is an innovative and effective solution to beat this challenge.

One common way to secure cyber-physical infrastructure is to isolate it in a confined room and restrict access [3]. However, it is only possible for server computers that allow remote access through

 doi:10.20944/preprints202303.0183.v1

computer networks. However, it is impossible to do it for desktops and laptops of the office desks. Organizations maintain security guards and keep the entrances locked during non-office hours. Many organizations maintain Close Circuit Television Camera (CCTV) and monitor everything from the control room [4]. Whether secured by guards or monitored from a control room through CCTV, it requires human involvement and undivided attention. It is beyond human capability to monitor the security status with maximum attention level because the average attention span of adults is 20 minutes [5]. It is a significant vulnerability in cyber-physical security. Applying Artificial Intelligence (AI)-driven solutions is a potential way to overcome this vulnerability [6]. The literature review shows the effective application of AI, including in healthcare [7], robotics [8], microbiology [9], image segmentation [10], and road construction [11]. The HAR is a subbranch of AI that has been applied in the proposed methodology to strengthen cyber-physical system security.

The proposed HAR-CPS algorithm uses a combination of GoogleNet [12] and BiLSTM [13] networks. The BiLSTM network learns from the features extracted by GoogleNet and later automatically recognize the activities it is trained to classify. Depending on the level of suspicious activities, the proposed HAR-CPS generates an alarm to alert the responsible authorities. This paper also focuses on the security of the proposed security system. That is why the entire system is deployed in the cloud so that the intruders fail to attack the proposed security system physically. A USB camera connected to an IoT device to transmit the video to the cloud is the only cyber-physical component of the proposed system. The core contributions of the proposed system are:

- Development and training of GoogleNet-BiLSTM hybrid network to classify designated human activities from video with an average accuracy of 73.15%.
- Creative design of the cyber-physical security system using IoT and cloud computing to ensure the cyber-physical security of the proposed security system.
- Formulation of the novel HAR-CPS algorithm to use GoogleNet-BiLSTM hybrid network to ensure security.
- Application of Machine Vision at the Edge (Mez) to minimize the cloud resource for cost minimization.

The rest of the paper has been organized into five sections. The second section contains the literature review. The methodology has been presented in the third section of this paper. The methodology is further divided into two more subsections - Dataset and Network architecture. The fourth section of this paper demonstrates the experimental results and performance evaluation. Finally, the paper has concluded in the fifth section.

2. Literature Review

According to the A. Ray et al. Human Activity Recognition (HAR) is a vibrant research field [14]. The recent advancements in this research domain demonstrate outstanding performances of the Convolutional Neural Network (CNN)-based approaches [15]. The commercial application of HAR technology is visible in different sectors, including the healthcare sector, fitness tracking, smart homes, smart surveillance & security, and sports analysis [16]. The proposed methodology of this paper is an application of HAR in cyber-physical security. The application of HAR technology in security is not new. L. P. O Paula et al. developed a front door security system using human activity recognition-based approach [17]. It strengthens the security at the front door by alerting respected authorities if violent activities are detected. The concepts of the proposed paper align with this paper. However, the HAR-CPS algorithm explores the potential of applying HAR in cyber-physical security.

Research conducted by B. Sarp et al. used a Raspberry Pi-based security system similar to the proposed methodology [18]. However, there is no artificial intelligence applied in their approach. It is a video and audio transmission system that allows users to see outdoor activities and maintain verbal communication. The proposed HAR-CPS algorithm is much more advanced. It uses a sophisticated GoogleNet-BiLSTM network to automatically classify the activities and notify the authority if there is

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 March 2023 doi:10.20944/preprints202303.0183.v1

any threat to cyber-physical security. The security system developed by R. C. Aldawira et al. has an innovative application of IoT, a motion sensor, and a touch sensor [19]. Despite the scope of applying HAR technology, most of the research focuses on video surveillance, and a simple sensor-based approach [20–22]. Compared to these papers, the proposed HAR-CSP algorithm is more advanced and effective than most of the state-of-the-art applications of HAR in securing cyber-physical systems.

M. Kong et al. have developed a real-time video surveillance system that addresses network latency challenges for real-time video communication [23]. Similar challenges have been faced in the edge computing-enabled video segmentation research conducted by S. Wan [24]. Transmitting video in real-time requires high bandwidth and is sensitive to time delay. A significant amount of time delay caused by latency interrupts the frame sequence [25]. Moreover, video processing requires high cloud resources, which increases the expenditure. According to M. Darwich, cost minimization for video processing provided through cloud services is essential [26]. Real-time video transmission through latency-sensitive networks and video processing in the cloud are two challenges the proposed methodology face as well. A. George et al. developed an effective communication technology for real-time video transmission through a latency-sensitive network while maintaining acceptable quality using Machine vision at the IoT Edge (Mez) [27]. The proposed methodology uses the Mez technology to manage the latency sensitivity and cloud resource usage for video processing.

Video analysis and its applications in intelligent surveillance, autonomous vehicles, video analysis, video retrieval, and entertainment rely heavily on computer vision-based human activity recognition [28]. This paper's analysis agrees with both observation and technique of the proposed methodology. While designing a cyber-physical system security algorithm, it is best to focus on combining computer vision and machine learning. A temporary posed-based human action recognition system was created by V. Mazzia1 et al. [29]. In a test with 227,000 parameters, it obtained 90.86 percent accuracy. While the paper's precision is impressive, the high computational cost renders it unsuitable for developing a cheap security system. A DCNN-based architecture using depth vision guided by Q. Wen et al. obtained a promising 93.89 percent accuracy [30]. To train robots on video datasets, this strategy overcomes the difficulty of collecting and classifying large amounts of data. The Microsoft Kinect camera is required for it, which is not cost-effective. Compared to these approaches, the proposed HAR-CPS algorithm is computationally simple and less expensive, yet a high-performing solution to cyber-physical system security [31].

3. Methodology

A GoogleNet-BiLSTM hybrid network is employed as the classifier in the proposed HAR-CPS algorithm. A video dataset is necessary for this type of hybrid network. In this section, we explained the HAR-CPS algorithm, along with the video dataset selection criteria, dataset processing, network design, the HAR-CPS method's operating principle, and mathematical interpretations. Figure 1 provides a visual summary of the proposed approach.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 March 2023 doi:10.20944/preprints202303.0183.v1

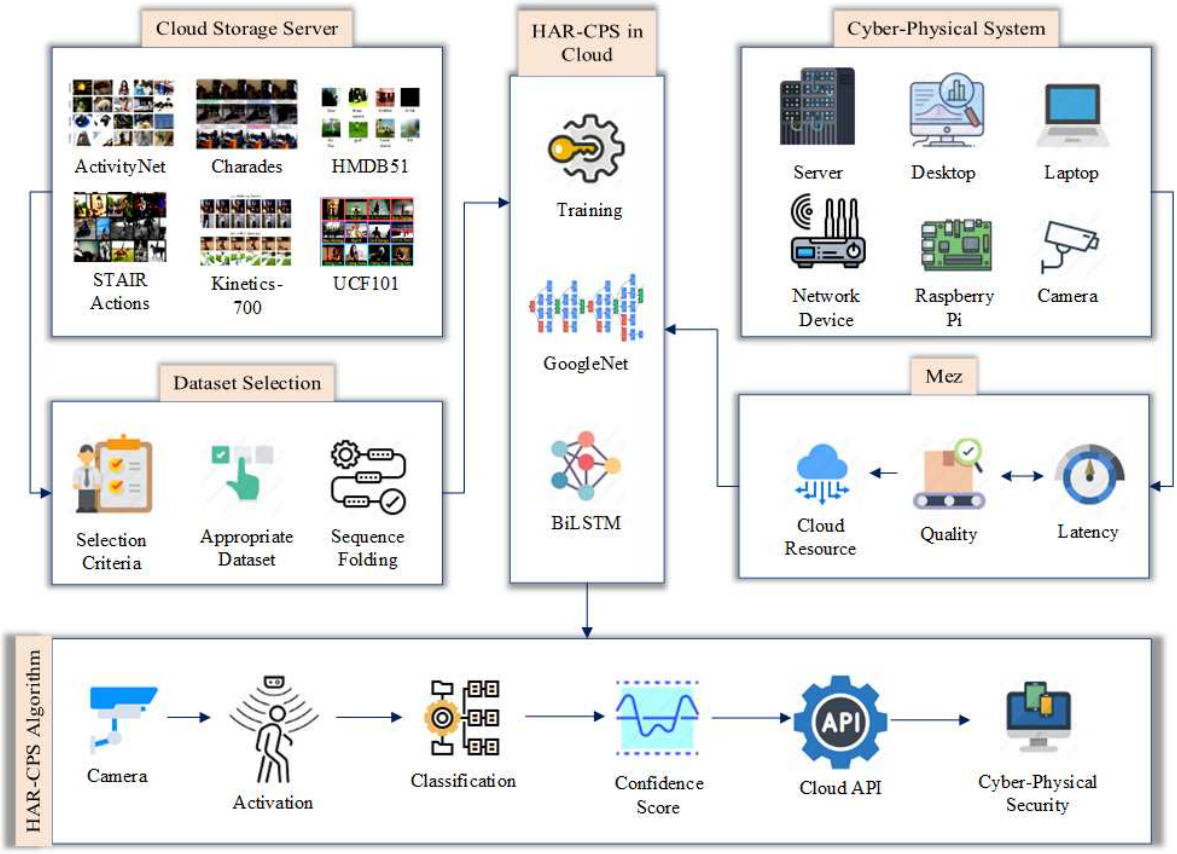

Figure 1. The overview of the proposed methodology

3.1. Dataset Selection

The proposed HAR-CPS algorithm is a human activity recognition-based approach. There are multiple Human Activity Recognition (HAR) datasets. This research has studied and analyzed the most widely used HAR datasets. These datasets are listed in Table 1 [28]. Each dataset is rich enough to train a CNN to recognize human activities. However, the purpose of this research is to recognize activities that are considered threats to the security of cyber-physical systems.

Table 1. Human Activity Recognition (HAR) dataset description

Dataset	Categories	Videos	Description
ActivityNet [32]	200	21,313	Activities conducted on a daily, social, and domestic basis, including games and workouts
Charades [33]	157	66,493	Routine chores performed within the house, such as refilling glasses and folding towels, etc.
HMDB51 [34]	51	5,100	Movement of the body and face, as well as contact with objects, are all included
Kinetics-700 [35]	700	530,336	Interactions involving a single person as well as those involving many people
STAIR Actions [36]	100	109,478	Frequent indoor activities in the house, workplace, bathroom, kitchen, item handling, etc.
UCF101 [37]	101	13,320	Interactions between humans and other objects, movements of the body that do not include other objects, and the utilization of various instruments.

Usually, large-scale cyber-physical systems are kept in confined rooms with limited access. Trained security personnel check the credentials of anyone who wants to access the cyber-physical systems. The proposed HAR-CPS algorithm aims to keep the physical computing infrastructure safe and monitor

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 March 2023 doi:10.20944/preprints202303.0183.v1

security breaches as real human security guards. Anyone to access the cyber-physical system without proper authorization and keys to unlock the doors will apply physical force to open the door. The attacker may punch the door to break it. Someone may try to break the door by kicking or hitting it. Pushing the door is another physical force someone may use to break it. Instead of physical force, intruders may carry weapons to gain access to cyber-physical systems. We have selected five activities listed in Table 2 from this observation. These five activities are our core dataset selection criteria.

Table 2. Description of the incidents and class names

Serial	Incident	Class
1	Trying to break the door by punching	Punch
2	Trying to kick open the door	Kick
3	Hitting on the doorknob to break it	Hit
4	Showing up in front of the door with weapon	Weapon
4	Pushing the door to open it forcefully	Push

According to our inspection, the HMDB51 dataset contains the target categories mentioned in Table 2. This dataset has a total of 47 categories of videos. The five selected activities are a subset of these 47 categories. That is why HMDB51 is the selected dataset for this experiment. The video clips of the HMDB51 dataset are realistic and original footage. There are no animation or made-up clips. That is why these videos do not require additional filtering and feature enhancement.

3.2. The Hybrid Network Architecture

The proposed CPS algorithm combines GoogleNet and a Long Short Term Memory (LSTM) network. GoogleNet is used to extract the features from the dataset. The LSTM network uses those features to recognize the activities in real time.

3.2.1. Sequence Folding

Detection, the suspicious activities in real-time are crucial in cyber-physical security. A grayscale video stream at 30 FPS contains more than 9000 frames in a 5 minutes video. At the same rate, 24-hour video footage contains 2.6×10^6 frames. The frames will be 3 times more if the color video is streamed. Extracting features directly from the video is impractical because of this large number of frames. It introduces very high latency. As a result, the system fails to detect suspicious activities in real time. We have used the sequence folding method defined by equation 1 to convert the video sequence into a separate set of images which is defined by equation 1.

$$\sum_{i=1}^{N} I(m_i, n_i) = \sum_{t=1}^{T} f_r((m_t, n_t), t) \tag{1}$$

The $f_r((m_t, n_t), t)$ in equation 1 is time-dependent frame. This time-dependent frame is converted into time-independent individual images expressed by$I(m_i, n_i)$. These frames are sent to the cloud server. The time-independent frames minimize the latency.

3.2.2. Feature Extractor Network in Cloud

Feature extraction from image frames is computationally expensive. The resource-constrained IoT devices are not suitable for it. We used a GoogleNet for feature extraction. Google Cloud has the GoogleNet readily available, which is a pretrained network. However, the entire GoogleNet has not been used. It is a 22-layer deep Convolutional Neural Network (CNN). We used didn't use the last three layers. The 19th layer is an average pooling layer. The features are available at this layer. The extracted features are converted into a feature vector using Algorithm 1.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 March 2023 doi:10.20944/preprints202303.0183.v1

Algorithm 1 Constructing Feature Vector

Input: GoogleNet, G_N; Frame, F
Output: Feature Vector, F_s;
Initiate: Allocate Virtual Machine, VM;
Start
$L_s \leftarrow VM(Size(Layers(1, G_N)))$
$L_s \leftarrow VM(Convert(L_s, F_s))$
for $i \leftarrow 1 : F$ **do**
    $Feature \leftarrow VM(pooling(F))$
    $F_s \leftarrow VM(Concat(Feature))$
end for
$VM(save(F_s))$
end

The Algorithm 1 initializes the Virtual Machine (VM) in the cloud to extract features from the images. It takes $475ms$ to initiate the virtual resources and an additional $711ms$ time to extract features per frame. It totally takes $1.19seconds$ to extra features from a one-minute video. The $1.19seconds$ time delay is considered as real-time.

3.2.3. GoogleNet-BiLSTM Hybridization

The BiLSTM network is ideal for classifying sequential data, and GoogleNet is optimally designed to extract distinguishable features from images. The hybridization of these two different networks develops a system efficient in feature extraction and sequential data classification. The GoogleNet-BiLSTM hybridization has been developed and studied from this observation, illustrated in Figure 2. The BiLSTM network in the experimental setup receives the video features from GoogleNet's average pooling layer. These features are passed to the BiLSTM layer. The responses from this layer are concatenated. These concatenated responses are sent to the dense layer. It follows a fully connected network architecture and a Softmax layer for classification. The classification layer has five output nodes. Each node produces a confidence score, representing the probability of being a certain class.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 March 2023 doi:10.20944/preprints202303.0183.v1

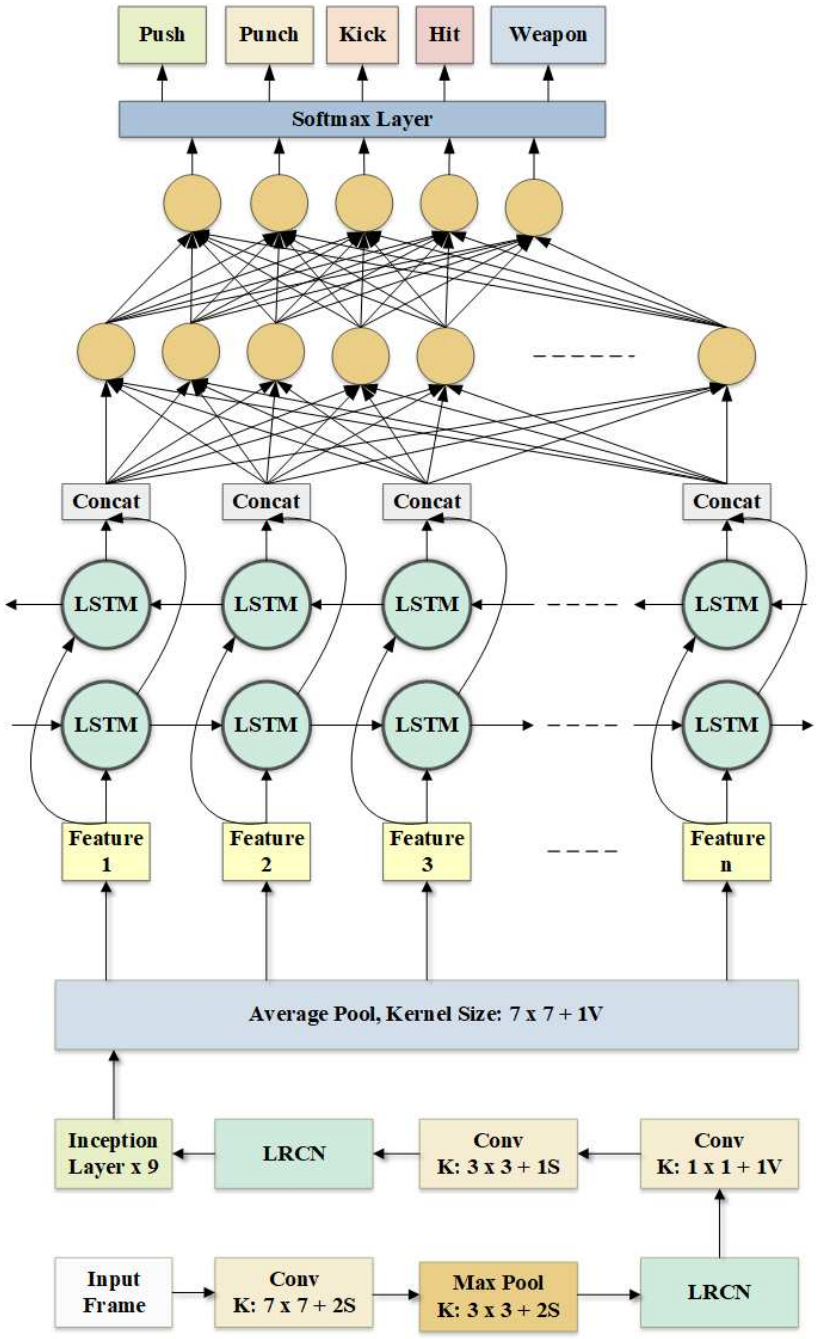

Figure 2. The GoogleNet-BiLSTM Hybridization

3.2.4. Training the Hybrid Network

The BiLSTM network has been trained with the features extracted from GoogleNet. The dataset has been split into training, testing, and validation dataset with a ratio of 70:15:15. The training dataset is used to train the network. The validation dataset has been used to validate the learning progress during the training. The testing dataset has been kept separated and untouched during the training period. It has been used to test the performance of the trained hybrid system during experimental analysis. Instead of using the entire dataset simultaneously, we used batch normalization with a mini-batch of size 16. During every iteration, the video clips are internally shuffled within the mini-batch.

Learning algorithms play a vital role in the collective performance of machine learning models. In this experiment, three widely used learning algorithms for deep neural networks have been

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 March 2023 doi:10.20944/preprints202303.0183.v1

studied. And they are Adaptive Gradient Algorithm (AdaGrad) [38], Root Mean Squared Propagation (RMSProp) [39], and Adaptive Moment Estimation (ADAM) [40]. These learning algorithms are expressed using equations 2, 3, and 4, respectively.

$$\omega_i^{(t+1)} = \omega_i^t - \frac{\eta}{\sqrt{\sum_{\tau=1}^{t} g_{\tau,i}^2}} g_{t,i} \tag{2}$$

$$\omega_i^{(t+1)} = \omega_i^t - \frac{\eta}{\sqrt{(v_t) + \epsilon}} \Delta_t \tag{3}$$

$$\omega_i^{(t+1)} = \omega_i^t - m_t \left(\frac{\alpha}{\sqrt{v_t} + \epsilon} \right) \tag{4}$$

The learning algorithms adjust the weights of the hidden nodes of deep neural networks. The more efficient this process is, the better the performance of the trained network becomes. We experimented with all three of the aforementioned algorithms and analyzed the performance using a validation loss curve illustrated in Figure 3.

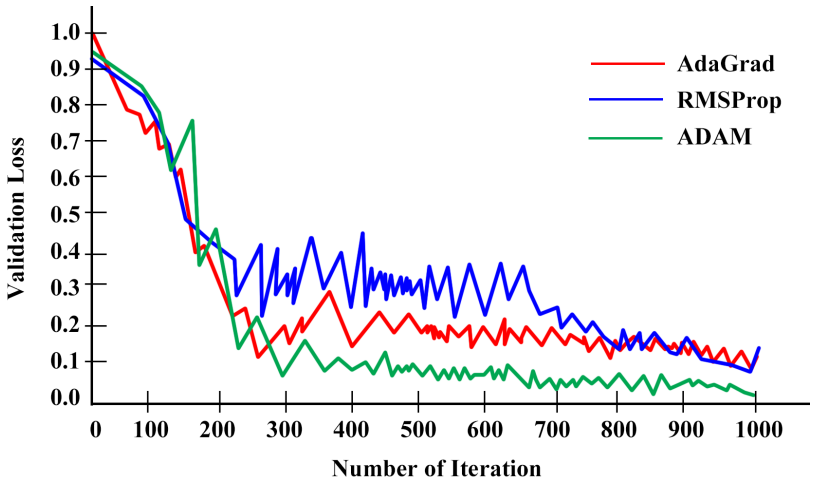

Figure 3. The optimization algorithm selection.

The validation loss curve shows that the AdaGrad learning algorithm reduces the validation loss to 250 iterations. However, there are lots of turbulence between 250 to 700 iterations. After that, the validation loss reduces again. Compared to this, RMSProp's performance is much better than AdaGrad's. However, the characteristics of the validation loss curve are almost similar. According to the experimental analysis in Figure 3, the ADAM is the best-performing learning algorithm. That is why ADAM has been used as the learning algorithm in this research. The proposed network has been trained with 1000 iterations and 568 epochs. The learning progress is illustrated in Figure 4.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 March 2023 doi:10.20944/preprints202303.0183.v1

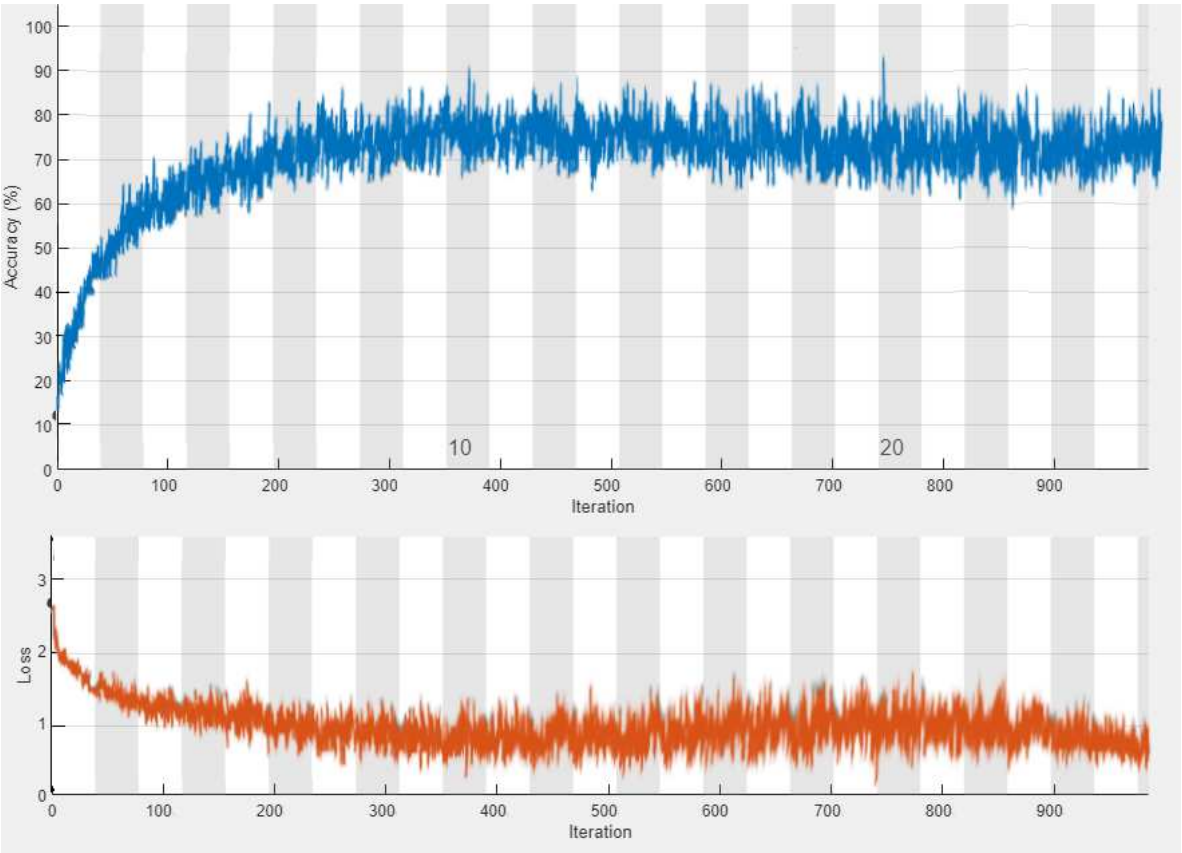

Figure 4. The learning curve with validation training accuracy and validation loss

It takes 342 minutes and 19 seconds to complete the training. It has been observed that the accuracy of the validation data increases sharply, and validation loss falls sharply till 200^{th} iterations. After that, the slope is negligible, and the learning curve maintains smooth progress. It ends with 72.48% validation accuracy. The initial learning rate is 0.001, and the final learning rate is 0.0001. The dynamic learning rate has been used in this experiment which adjusts itself depending on the accuracy and loss.

3.2.5. HAR-CPS Algorithm

The proposed innovative HAR-CPS algorithm, presented as Algorithm 2, uses the trained GoogleNet-BiLSTM hybrid network to classify the target categories. It runs in a virtual machine provisioned through a pay-as-you-go payment method. It is more efficient to reduce the computational resource to minimize the cost. The proposed algorithm has been designed to minimize the cost. Human activity recognition is the most computationally expensive process. The algorithm calls GoogleNet-BiLSTM hybrid network only when necessary. For the rest of the time, it performs simple linear 2D subtraction. As a result, the cost becomes minimum.

The Algorithm 2 applies GoogleNet-BiLSTM to recognize human activities only when two successive frames have more than 70% dissimilarity. Calculating the dissimilarity requires insignificant computing resources. For a 2MP camera running 24×7 with a frame rate of 30FPS, the monthly cost of frame difference calculation is less than $4.29. Once two successive frames have more than 70% difference, the proposed HAR-CPS algorithm passes the frame to the GoogleNet-BiLSTM network. It predicts the human activity on the video stream and returns a confidence score. If the confidence score is more than 80%, the alert is generated through a security API.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 March 2023 doi:10.20944/preprints202303.0183.v1

Algorithm 2 The HAR-CPS Algorithm

Input: CCTV Video Stream, v_s; HLS Request, H_l
Initiate: Allocate Cloud Resource;
Output: Alert, a;
Start
$k \leftarrow 0$
$F[k] \leftarrow read(v_s)$
while $v_s = True$ **do**
  $i \leftarrow i + 1$
  Accept HLS Request
  $F[i] \leftarrow read(v_s)$
  $d \leftarrow difference(F[i-1], F[i])$
  if $d \geq 0.70$ **then**
    $[p, s] \leftarrow GoogleNetBiLSTM[F[i]]$
    if $s \geq 0.80$ **then**
      $a \leftarrow class(p)$
      $SecurityAPI(a)$
    end if
  else
    $NoAction$
  end if
end while
end

3.3. Latency & Cloud Resource Optimization using Mez

The original Mez architecture was built to link several IoT camera nodes simultaneously. The Edge server is linked to it through a wireless network [27]. In the suggested setup, only one camera is linked to a Raspberry Pi 4. Unlike the original Mez system, the proposed system communicates with the cloud server over a licensed 4G spectrum. As a result, a modified Mez architecture, as shown in Figure 5, was adopted in this experiment. This architecture includes a 4G network sensor to check network quality. It exchanges data with the Network Latency Controller.

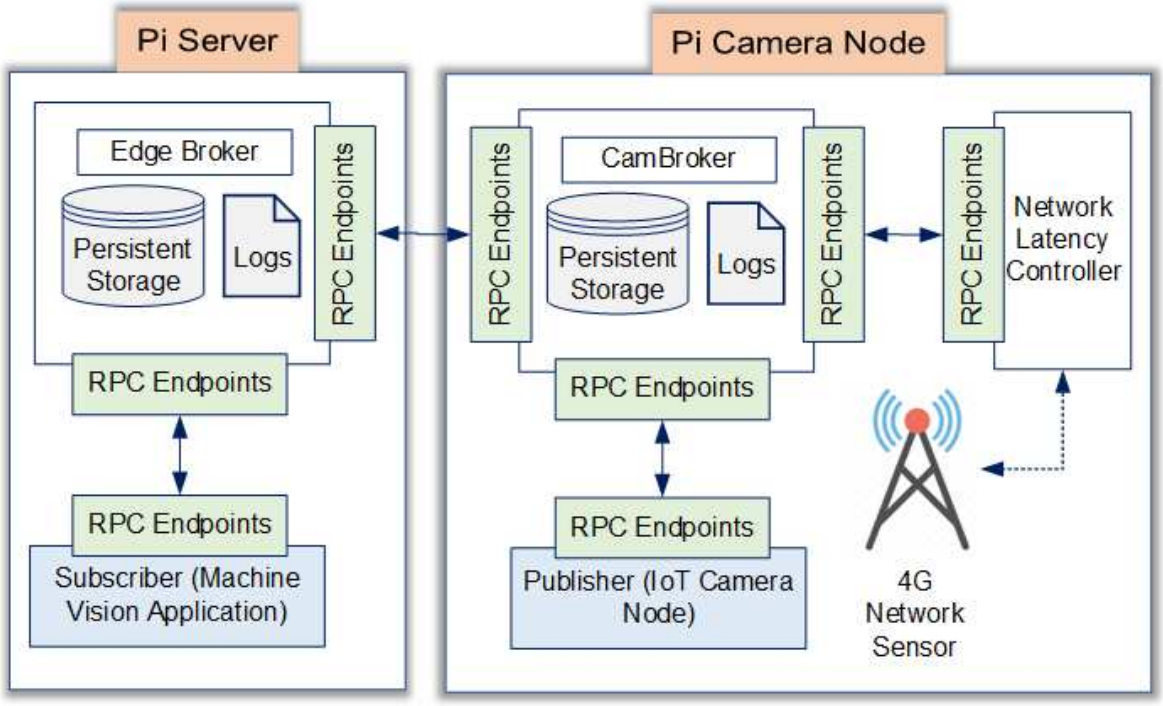

Figure 5. The Mez architecture

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 March 2023 doi:10.20944/preprints202303.0183.v1

3.3.1. Latency VS. Quality Trade-off

The suggested system uses Mez technology's latency VS. quality trade-off capabilities. The frame quality may be adjusted using five different knob settings depending on the application precision requirements. Table 3 lists the possible knob settings, their functions, the influence on frame size reduction, and the application scopes.

Table 3. The knob configuration and effects

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%	Countour-based approach
5	Frame Differincing	40%	Linear frame difference based method

3.3.2. Cloud Resource Optimization

The proposed HAR-CPS algorithm optimizes cloud resource usage using the Mez [27] technology. The empirical analysis shows that keeping the first knob setting listed in Table 3 at 940×528 resolution reduces the frame size by 8% lowering the cloud resource usage for video processing. The grayscale colorspace has been used, which reduces the frame size by 11%. Although Table 3 shows that blurring reduces the frame size, the proposed methodology does not use this knob. It has been observed that blurring the video downgrades the feature quality extracted by GooleNet. However, the Artifact Removal and Frame Difference knobs have been used, and they reduced the frame size by 14% and 16%, respectively. After applying the Mez technology, the average frame size reduction becomes 49%. As a result, cloud resource usage is reduced by almost 50%.

4. Results and Performance Evaluation

The proposed human activity recognition-based cyber-physical security algorithm is a deep learning-based approach that runs on a cloud server. The performance of the system has been evaluated from two different perspectives. First, the first of the proposed GoogleNet-BiLSTM hybrid network has been evaluated. After that, the performance of the cloud system was studied.

Table 4. Performance comparison of the proposed system with different models and video lengths

Model Name	Frame Sequence							
	30 Seconds Clips				60 Seconds Clips			
	Accuracy	Precision	Recall	F-1 Score	Accuracy	Precision	Recall	F-1 Score
BiLSTM	70.45%	68.41%	65.41%	62.40%	72.45%	69.74%	68.41%	58.41%
CNN	63.47%	65.71%	63.91%	60.84%	65.44%	69.71%	62.48%	57.94%
MLP	65.71%	62.78%	65.46%	61.75%	66.78%	65.17%	65.17%	55.17%
LSTM	67.40%	64.71%	66.34%	65.37%	68.41%	62.47%	66.34%	62.78%
Proposed Model	74.17%	72.85%	67.46%	66.74%	74.79%	73.01%	68.70%	67.41%

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 March 2023 doi:10.20944/preprints202303.0183.v1

4.0.1. Performance of GoogleNet-BiLSTM Hybrid Network

The performance of the proposed GoogleNet-BiLSTM hybrid network has been evaluated using state-of-the-art machine learning performance evaluation metrics. The literature review shows that machine learning-based image classification where CNN or LSTM networks are utilized use accuracy, sensitivity, specificity, False Positive Rate (FPR), and False Negative Rate (FNR) [7]. The mathematical definitions of these evaluation metrics are listed in Table 5. These values are calculated from the True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN), which are obtained from the confusion matrix illustrated in Figure 6.

Table 5. The evaluation metrics used in this research

Evaluation Metrics	Mathematical Expression	Role
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$	Classification accuracy
Sensitivity	$\frac{TP}{TP+FN}$	Correct identification of actual positive cases
Specificity	$\frac{TN}{TN+FP}$	True negative rate
False Positive Rate	$1 - Specificity$	Type I error
False Negative Rate	$1 - Sensitivity$	Type II error

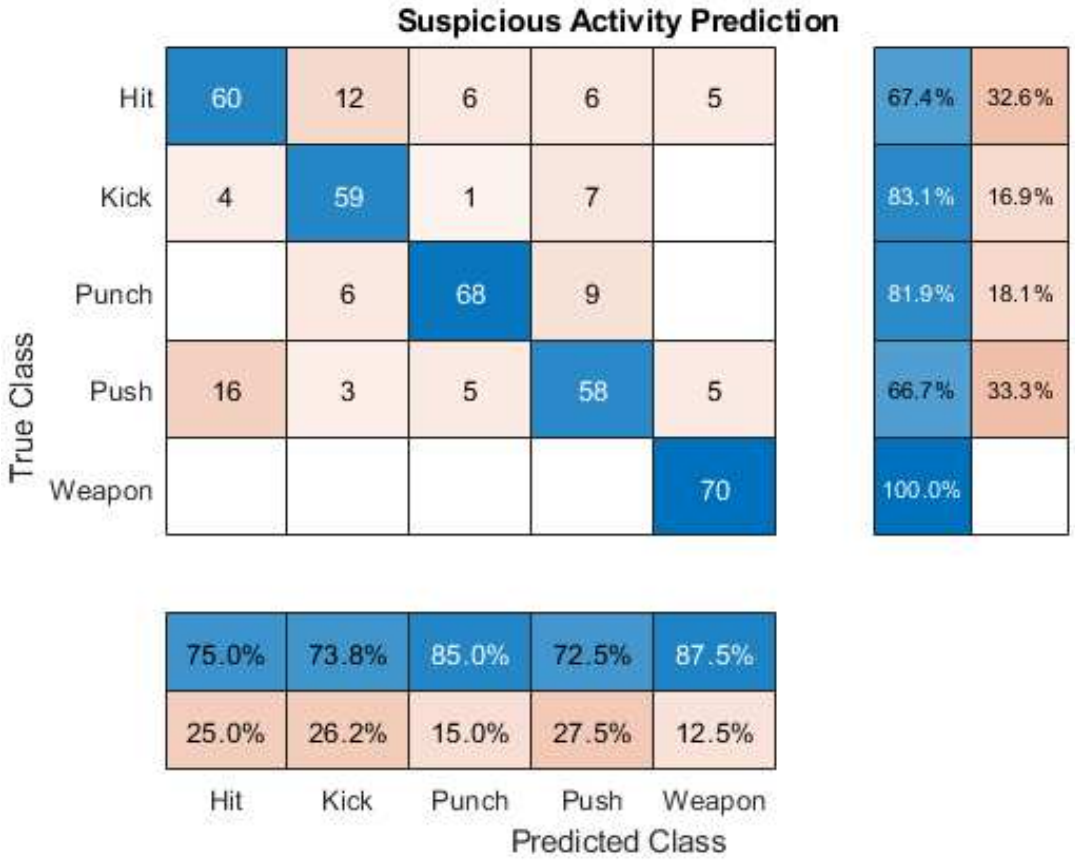

Figure 6. The confusion matrix for performance analysis

The performance of the proposed GoogleNet-BiLSTM network in terms of the state-of-the-art machine learning evaluated metrics listed in Table 5 is listed in Table 6 [7]. The experimental result shows that the proposed system best classifies the 'kick' category. The average classification accuracy

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 March 2023 doi:10.20944/preprints202303.0183.v1

is 73.15%. The average sensitivity, specificity, false positive rate, and false negative rate are 71.52%, 72.22%, 28.48%, and 27.78%, respectively.

Table 6. Classification performance of the GoogleNet-BiLSTM Network

Activity	Accuracy	Sensitivity	Specificity	FPR	FNR
Hit	73.10%	70.0%	62.2%	30.0%	37.8%
Kick	76.78%	61.3%	80.3%	38.7%	19.7%
Punch	71.47%	80.0%	75.3%	20.0%	24.7%
Push	68.63%	72.5%	66.7%	27.5%	33.3%
Weapon	75.79%	73.8%	76.60 %	26.2%	23.4%

4.0.2. Performance Comparison

The performance of the proposed system has been compared with four different models. These models are BiLSTM, CNN [41], MLP [42], and LSTM [43]. The experimenting dataset has different lengths of videos. We categorized them into 30 seconds and 60 seconds video clips. This experiment has been conducted to understand the effect of the proposed system on video clips with different duration. The result of the experiment has been listed in Table 4, demonstrating that the proposed system outperforms other similar approaches.

4.0.3. Resource Optimization Performance

The proposed GoogleNet-BiLSTM hybrid network runs in the cloud server, which handles the video stream from the proposed system[44]. Cloud resource optimization is a major contribution of the proposed methodology. A pay-as-you-go payment scheme has been used to implement the HAR-CPS algorithm. That means the expenditure increases with resource usage. The cloud resource optimization statistics over 60 minutes which has been averaged every 10 minutes are listed in Table 7. The statistical data shows that the optimization scheme used in this paper is most effective in primary memory usage reduction. It reduces the primary memory consumption by 64.44%. It has a positive effect on CPU usage as well. The proposed HAR-CPS system uses 0.45% less CPU after resource optimization. The average disk writing time is 0.12 MB/s after using the Mez, which is a 43.58% reduction.

Table 7. The cloud resource optimization statistics over 60 minutes

	Without Mez			With Mez		
Time	CPU (%)	Memory (MB)	Disk (MB/s)	CPU (%)	Memory (MB)	Disk (MB/s)
10	0.2	151	0.10	0.1	37	0.13
20	0.8	155	0.20	0.5	47	0.13
30	1.1	90	0.10	0.4	57	0.13
40	1.2	78	0.30	0.1	36	0.13
50	0.7	120	0.30	0.3	50	0.07
60	0.7	140	0.30	0.6	34	0.13

5. Limitation & Future Scope

The experimental results and performance evaluation demonstrate the acceptability of the proposed HAR-CPS algorithm to strengthen the security of cyber-physical systems. Despite the impressive performance, it has several limitations, which have been discussed in this section. However, instead of considering them as limitations, these have been considered as the future scope of this research. These limitations are:

5.1. Limited Number of Actions

The proposed algorithm effectively classifies five human actions that are potential threats to cyber-physical system security. However, more actions may be considered a security risk that this

 doi:10.20944/preprints202303.0183.v1

paper has not considered. The limited number of actions is a significant limitation of this research. The GoogleNet-BiLSTM hybrid network has the potential to learn to classify hundreds of different types of actions. It requires datasets with more categories. The subsequent version of the proposed HAR-CPS will be trained to categorize more human activities to ensure more rigorous cyber-physical security.

5.2. Camera-Subject Angle Sensitivity

The proposed system's accuracy is sensitive to the viewing angle between the subject and the camera. The intruders must be within the 40 to 60 degrees viewing angle. Although the camera is placed on maintaining this particular viewing angle, it is still considered a weakness of the system. A geometrical image transformation algorithm is a potential solution to reduce the camera-subject angle sensitivity. The subsequent research on the proposed HAR-CPS algorithm will explore this opportunity.

5.3. Security of the HAR-CPS Device

A significant portion of the proposed HAR-CPS algorithm runs on the cloud server. As a result, it is secured from cyber-physical attacks. However, imaging and IoT devices are kept on the premises and vulnerable to cyber-physical attacks. A creative camouflage deployment model is a potential solution to this problem, opening new research opportunities.

It is beyond the scope of any approach to ensure 100% security. There are always weaknesses in security systems. The proposed HAR-CPS system is no different. It is effective in strengthening cyber-physical security within its application domain. The limitations of the proposed system pave the path to conducting more research in this domain and developing a better version of the HAR-CPS algorithm.

6. Conclusion

Cyber-Physical security is the protection of critical infrastructure systems that are integrated with computer networks and software. Both both physical and digital components are affected in case of a cyber-physical security breach. Firewalls, intrusion detection systems, frequent vulnerability assessments, and other forms of cyber and physical security, such as access control and surveillance, must be put in place to ensure the safety of these systems. However, implementing cyber-physical system surveillance and security is more challenging than software-based cybersecurity. The Human Activity Recognition-based Cyber-Physical Security (HAR-CPS) algorithm beats this challenge with flying colors. It reduces the necessity of human intervention from cyber-physical security surveillance and automatically recognizes suspicious activities with an average accuracy of 73.15%. The innovative GoogleNet-BiLSTM network-based classifier and the algorithm run on the cloud server, away from the cyber-physical system. As a result, the proposed system remains secured when the cyber-physical system is under attack. The effective application of the Mez technology automatically adjusts the video quality to tolerate the latency sensitivity and prevents real-time video transmission interruption. It also reduces the frame size, which optimizes the cloud server expenditure. That is why the innovative HAR-CPS algorithm strengthens cyber-physical security with $4.29 only per month.

References

1. Duo, W.; Zhou, M.; Abusorrah, A. A survey of cyber attacks on cyber physical systems: Recent advances and challenges. *IEEE/CAA Journal of Automatica Sinica* **2022**, *9*, 784–800.
2. Zhao, Z.; Xu, Y. Performance based attack detection and security analysis for cyber-physical systems. *International Journal of Robust and Nonlinear Control* **2023**, *33*, 3267–3284.
3. Hammoudeh, M.; Epiphaniou, G.; Pinto, P. Cyber-Physical Systems: Security Threats and Countermeasures, 2023.
4. De Pascale, D.; Sangiovanni, M.; Cascavilla, G.; Tamburri, D.A.; Van Den Heuvel, W.J. Securing Cyber-Physical Spaces with Hybrid Analytics: Vision and Reference Architecture. Computer Security.

ESORICS 2022 International Workshops: CyberICPS 2022, SECPRE 2022, SPOSE 2022, CPS4CIP 2022, CDT&SECOMANE 2022, EIS 2022, and SecAssure 2022, Copenhagen, Denmark, September 26–30, 2022, Revised Selected Papers. Springer, 2023, pp. 398–408.

5. Jadhao, A.; Bagade, A.; Taware, G.; Bhonde, M. Effect of background color perception on attention span and short-term memory in normal students. *National Journal of Physiology, Pharmacy and Pharmacology* **2020**, *10*, 981–984.
6. Del Giudice, M.; Scuotto, V.; Orlando, B.; Mustilli, M. Toward the human–centered approach. A revised model of individual acceptance of AI. *Human Resource Management Review* **2023**, *33*, 100856.
7. Faruqui, N.; Yousuf, M.A.; Whaiduzzaman, M.; Azad, A.; Barros, A.; Moni, M.A. LungNet: A hybrid deep-CNN model for lung cancer diagnosis using CT and wearable sensor-based medical IoT data. *Computers in Biology and Medicine* **2021**, *139*, 104961.
8. Chakraborty, P.; Yousuf, M.A.; Zahidur Rahman, M.; Faruqui, N. How can a robot calculate the level of visual focus of human's attention. Proceedings of International Joint Conference on Computational Intelligence: IJCCI 2019. Springer, 2020, pp. 329–342.
9. Trivedi, S.; Patel, N.; Faruqui, N. Bacterial Strain Classification using Convolutional Neural Network for Automatic Bacterial Disease Diagnosis. 2023 13th International Conference on Cloud Computing, Data Science & Engineering (Confluence). IEEE, 2023, pp. 325–332.
10. Trivedi, S.; Patel, N.; Faruqui, N. NDNN based U-Net: An Innovative 3D Brain Tumor Segmentation Method. 2022 IEEE 13th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON). IEEE, 2022, pp. 0538–0546.
11. Arman, M.S.; Hasan, M.M.; Sadia, F.; Shakir, A.K.; Sarker, K.; Himu, F.A. Detection and classification of road damage using R-CNN and faster R-CNN: a deep learning approach. Cyber Security and Computer Science: Second EAI International Conference, ICONCS 2020, Dhaka, Bangladesh, February 15-16, 2020, Proceedings 2. Springer, 2020, pp. 730–741.
12. Ibrahim, Y.; Wang, H.; Adam, K. Analyzing the reliability of convolutional neural networks on gpus: Googlenet as a case study. 2020 International Conference on Computing and Information Technology (ICCIT-1441). IEEE, 2020, pp. 1–6.
13. Wei, X.; Wu, J.; Ajayi, K.; Oyen, D. Visual descriptor extraction from patent figure captions: a case study of data efficiency between BiLSTM and transformer. Proceedings of the 22nd ACM/IEEE Joint Conference on Digital Libraries, 2022, pp. 1–5.
14. Ray, A.; Kolekar, M.H.; Balasubramanian, R.; Hafiane, A. Transfer Learning Enhanced Vision-based Human Activity Recognition: A Decade-long Analysis. *International Journal of Information Management Data Insights* **2023**, *3*, 100142.
15. Park, H.; Kim, N.; Lee, G.H.; Choi, J.K. MultiCNN-FilterLSTM: Resource-efficient sensor-based human activity recognition in IoT applications. *Future Generation Computer Systems* **2023**, *139*, 196–209.
16. Kulsoom, F.; Narejo, S.; Mehmood, Z.; Chaudhry, H.N.; Bashir, A.K.; others. A review of machine learning-based human activity recognition for diverse applications. *Neural Computing and Applications* **2022**, pp. 1–36.
17. Paula, L.P.O.; Faruqui, N.; Mahmud, I.; Whaiduzzaman, M.; Hawkinson, E.C.; Trivedi, S. A Novel Front Door Security (FDS) Algorithm using GoogleNet-BiLSTM Hybridization. *IEEE Access* **2023**.
18. Sarp, B.; Karalar, T. Real time smart door system for home security. *International Journal of Scientific Research in Information Systems and Engineering (IJSRISE)* **2015**, *1*, 121–123.
19. Aldawira, C.R.; Putra, H.W.; Hanafiah, N.; Surjarwo, S.; Wibisurya, A.; others. Door security system for home monitoring based on ESp32. *Procedia Computer Science* **2019**, *157*, 673–682.
20. Sanjay Satam, S.; El-Ocla, H. Home Security System Using Wireless Sensors Network. *Wireless Personal Communications* **2022**, pp. 1–17.
21. Banerjee, P.; Datta, P.; Pal, S.; Chakraborty, S.; Roy, A.; Poddar, S.; Dhali, S.; Ghosh, A. Home Security System Using RaspberryPi. In *Advanced Energy and Control Systems*; Springer, 2022; pp. 167–176.
22. Tao, J.; Wu, H.; Deng, S.; Qi, Z. Overview of Intelligent Home Security and Early Warning System based on Internet of Things Technology. *International Core Journal of Engineering* **2022**, *8*, 727–732.
23. Kong, M.; Guo, Y.; Alkhazragi, O.; Sait, M.; Kang, C.H.; Ng, T.K.; Ooi, B.S. Real-time optical-wireless video surveillance system for high visual-fidelity underwater monitoring. *IEEE Photonics Journal* **2022**, *14*, 1–9.

24. Wan, S.; Ding, S.; Chen, C. Edge computing enabled video segmentation for real-time traffic monitoring in internet of vehicles. *Pattern Recognition* **2022**, *121*, 108146.
25. Ujikawa, H.; Okamoto, Y.; Sakai, Y.; Shimada, T.; Yoshida, T. Time distancing to avoid network microbursts from drones' high-definition video streams. *IEICE Communications Express* **2023**, p. 2022XBL0184.
26. Darwich, M.; Ismail, Y.; Darwich, T.; Bayoumi, M. Cost Minimization of Cloud Services for On-Demand Video Streaming. *SN Computer Science* **2022**, *3*, 226.
27. George, A.; Ravindran, A.; Mendieta, M.; Tabkhi, H. Mez: An adaptive messaging system for latency-sensitive multi-camera machine vision at the iot edge. *IEEE Access* **2021**, *9*, 21457–21473.
28. Kong, Y.; Fu, Y. Human action recognition and prediction: A survey. *International Journal of Computer Vision* **2022**, *130*, 1366–1401.
29. Mazzia, V.; Angarano, S.; Salvetti, F.; Angelini, F.; Chiaberge, M. Action Transformer: A self-attention model for short-time pose-based human action recognition. *Pattern Recognition* **2022**, *124*, 108487.
30. Qi, W.; Wang, N.; Su, H.; Aliverti, A. DCNN based human activity recognition framework with depth vision guiding. *Neurocomputing* **2022**, *486*, 261–271.
31. Hesse, N.; Baumgartner, S.; Gut, A.; Van Hedel, H.J. Concurrent Validity of a Custom Method for Markerless 3D Full-Body Motion Tracking of Children and Young Adults based on a Single RGB-D Camera. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **2023**.
32. Caba Heilbron, F.; Escorcia, V.; Ghanem, B.; Carlos Niebles, J. Activitynet: A large-scale video benchmark for human activity understanding. Proceedings of the ieee conference on computer vision and pattern recognition, 2015, pp. 961–970.
33. Sigurdsson, G.A.; Gupta, A.; Schmid, C.; Farhadi, A.; Alahari, K. Charades-ego: A large-scale dataset of paired third and first person videos. *arXiv preprint arXiv:1804.09626* **2018**.
34. Sharma, V.; Gupta, M.; Pandey, A.K.; Mishra, D.; Kumar, A. A Review of Deep Learning-based Human Activity Recognition on Benchmark Video Datasets. *Applied Artificial Intelligence* **2022**, *36*, 2093705.
35. Carreira, J.; Noland, E.; Hillier, C.; Zisserman, A. A short note on the kinetics-700 human action dataset. *arXiv preprint arXiv:1907.06987* **2019**.
36. Yoshikawa, Y.; Lin, J.; Takeuchi, A. Stair actions: A video dataset of everyday home actions. *arXiv preprint arXiv:1804.04326* **2018**.
37. Soomro, K.; Zamir, A.R.; Shah, M. UCF101: A dataset of 101 human actions classes from videos in the wild. *arXiv preprint arXiv:1212.0402* **2012**.
38. Lydia, A.; Francis, S. Adagrad—an optimizer for stochastic gradient descent. *Int. J. Inf. Comput. Sci* **2019**, *6*, 566–568.
39. Turitsyn, S.K.; Schafer, T.; Mezentsev, V.K. Generalized root-mean-square momentum method to describe chirped return-to-zero signal propagation in dispersion-managed fiber links. *IEEE Photonics Technology Letters* **1999**, *11*, 203–205.
40. Newey, W.K. Adaptive estimation of regression models via moment restrictions. *Journal of Econometrics* **1988**, *38*, 301–339.
41. Albawi, S.; Mohammed, T.A.; Al-Zawi, S. Understanding of a convolutional neural network. 2017 international conference on engineering and technology (ICET). Ieee, 2017, pp. 1–6.
42. Riedmiller, M.; Lernen, A. Multi layer perceptron. *Machine Learning Lab Special Lecture, University of Freiburg* **2014**, pp. 7–24.
43. Bin, Y.; Yang, Y.; Shen, F.; Xu, X.; Shen, H.T. Bidirectional long-short term memory for video description. Proceedings of the 24th ACM international conference on Multimedia, 2016, pp. 436–440.
44. Hossen, R.; Whaiduzzaman, M.; Uddin, M.N.; Islam, M.J.; Faruqui, N.; Barros, A.; Sookhak, M.; Mahi, M.J.N. Bdps: An efficient spark-based big data processing scheme for cloud fog-iot orchestration. *Information* **2021**, *12*, 517.