IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 30 January 2023, accepted 10 February 2023, date of publication 23 February 2023, date of current version 28 February 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3248509

RESEARCH ARTICLE

A Novel Front Door Security (FDS) Algorithm Using GoogleNet-BiLSTM Hybridization

LUIZ PAULO OLIVEIRA PAULA[1], NURUZZAMAN FARUQUI[2], IMRAN MAHMUD[2], MD. WHAIDUZZAMAN[3], (Senior Member, IEEE), ERIC CHARLES HAWKINSON[4], AND SANDEEP TRIVEDI[5], (Senior Member, IEEE)

[1]Centro Universitário UniBTA, São Paulo 01310-300, Brazil
[2]Department of Software Engineering, Daffodil International University, Birulia 1216, Bangladesh
[3]School of Information Systems, Queensland University of Technology, Brisbane, QLD 4000, Australia
[4]Department of Global Tourism, Kyoto University of Foreign Studies, Kyoto 615-8558, Japan
[5]Deloitte Consulting LLP, Houston, TX 77002, USA

Corresponding author: Luiz Paulo Oliveira Paula (luizpaulo.oliveirapaula@gmail.com)

ABSTRACT Security has always been a significant concern since the dawn of human civilization. That is why we build houses to keep ourselves and our belongings safe. And we do not hesitate to spend a lot on front-door locks and install CCTV cameras to monitor security threats. This paper presents an innovative automatic Front Door Security (FDS) algorithm that uses Human Activity Recognition (HAR) to detect four different security threats at the front door from a real-time video feed with 73.18% accuracy. The activities are recognized using an innovative combination of GoogleNet-BiLSTM hybrid network. This network receives the video feed from the CCTV camera and classifies the activities. The proposed algorithm uses this classification to alert any attempts to break the door by kicking, punching, or hitting. Furthermore, the proposed FDS algorithm is effective in detecting gun violence at the front door, which further strengthens security. This Human Activity Recognition (HAR)-based novel FDS algorithm demonstrates the potential of ensuring better safety with 71.49% precision, 68.2% recall, and an F1-score of 0.65.

INDEX TERMS Intelligent surveillance, real-time security, deep learning, hybrid networks, sequence folding, video-frame feature vector.

I. INTRODUCTION

The application of Artificial Intelligence (AI) has upgraded different aspects of our lives. The banking [1], healthcare [2], education [3], agriculture [4], industrial automation [5], transportation [6], and many different sectors are leveraging the AI technology to reduce human intervention and automating services. The application of AI in cybersecurity and its effectiveness is worth attention as well [7]. However, the application of AI in physical security is still an under-explored area. This paper explores the effectiveness of applying AI in strengthening front-door security through an innovative algorithm named FDS. This paper is about an innovation to automate front-door security systems to replace the necessity of constant human intervention. The proposed algorithm mimics human-like intelligence to recognize gun violence and attempt to break the front door by hitting, kicking, or punching. And it alerts the residents like a loyal security guard.

In this paper, we created a hybrid network combining GoogleNet and BiLSTM network to introduce human-like intelligence in a front-door video surveillance system. The GoogleNet, a 22-layers deep Convolutional Neural Network (CNN), developed by a group of researchers of Google [8], can extract the image features the way the human visual cortex does [9]. And the Bidirectional Long Sort Term Memory (BiLSTM) can learn these features to classify YouTube videos [10], identify human sentiments [11], and recognize human activities [12]. That means a combination of these two networks exhibits the potential to extract video features from

The associate editor coordinating the review of this manuscript and approving it for publication was Seifedine Kadry.

video streams and identify human activities from them. This functionality is the heart of the FDS algorithm proposed in this paper.

We were under the impression of effortlessly building an automatic intelligent front-door security system. However, the reality is different, filled with challenges and difficulties. The challenges we faced, solutions we discovered, and the problems we overcame have been presented, analyzed, and discussed in this paper. Our findings, limitations, and solutions are worth recognition as novel contributions because it paves the path to ensure an automatic intelligent security system for everyone at an affordable cost. False alarm prevention is one of the challenges of intelligent surveillance systems [13]. The higher the accuracy of human activity detection, the lower the probability of having false alarms [14]. However, apart from accuracy, there are some other context-specific factors that govern the false alarm rate [15]. The methodology discussed in this paper recognizes human activities from live video feed with acceptable accuracy and solves the false alarm problem. The core contributions of this paper are listed as follows:

- An innovative hybrid network design for HAR-based security system development with an average accuracy of 73.18%.
- Development of a novel algorithm using CNN-BiLSTM combination in intelligent surveillance.
- A cost-effective solution with a nominal upfront cost without any subscription fee to make automatic and intelligent security affordable for everyone.

The rest of the paper has been organized into seven 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. The real-world implementation and its analysis have been presented in the fifth section. We have discussed the limitation and future scope of the proposed system in the sixth section. Finally, the paper has been concluded in the seventh section.

II. LITERATURE REVIEW

We have reviewed the recent and relevant literature on Human Activity Recognition (HAR)-based intelligent security systems. It shows the envious advancement in HAR and its application in various domains [16]. The performance of the HAR system seems to draw the lion's share of researchers' attention, leaving a research gap in its application in the security sector [17]. This paper focuses on applying HAR in intelligence surveillance to strengthen front door security. And the HAR technology is the main engine of this approach. That is why our literature review focuses on advancing HAR using machine learning and its application.

A. SMART APPLICATION FOR FRONT DOOR SECURITY

B. Sarp et al. used a Raspberry Pi-based video surveillance system to ensure front door security through two features - video feed and communication. In their system, the users can monitor the activities in front of the door remotely and also communicate with someone at the front door. They further connected the door through a cellular network to access the functionality in real-time through the internet [18]. While this approach effectively ensures front door security, it has a drawback. And the drawback is the necessity of manual inspection. The proposed FDS system does not require human intervention to monitor the front door security. It is a fully automatic system that identifies the activities of the individuals at the front door and alerts the homeowner if anything suspicious happens.

A home monitoring system based on ESP32, published by R. C. Aldawira et al., shows the application of IoT to ensure home security, including front door security. This system allows the users to monitor the activities happening inside remotely and outside the house and control the door lock. It also has a motion sensor to sense any motion and alert the users. Moreover, it has a touch sensor that is used to identify human touch on the door knob [19]. These multiple features make the home more secure. However, the system does not use human-like intelligence. Because of using motion and touch sensors, the rate of false alarms is high, and it requires manual adjustment. Compared to this approach, the proposed FDS is more advanced as it uses CNN and recognizes activities as the human visual cortex does [20]. IoT-based home security systems [21], edge computer-based security systems [22], and intelligent warning-based security systems [23] are the common approaches to enhance the security of home. The literature review demonstrates a research gap in front-door security using a convolutional neural network. The proposed FDS algorithm aims to abridge the gap and utilize CNNs to ensure human-guard-like security at the front door.

B. COMPUTER VISION-BASED HAR & APPLICATION

Computer vision-based human activity recognition is the dominating technology in video analysis and its application in intelligent surveillance, autonomous vehicle, video analysis, video retrieval, and entertainment [24]. The review presented in this paper aligns without observation and methodology. For a front-door security algorithm, a computer vision-based machine learning-centered approach is appropriate. V. Mazzia1 et al. developed a short-term posed-based human action recognition system. It achieved 90.86% accuracy with 227,000 parameters [25]. The accuracy of this paper is eye-catching, but the computational cost makes it expensive, which is not suitable for developing an affordable security system using this methodology.

A promising 93.89% accuracy was achieved by a DCNN-based framework with depth vision guided by Q. Wen et al. [26]. This methodology solves the extensive data collection and labeling challenge to train machines on

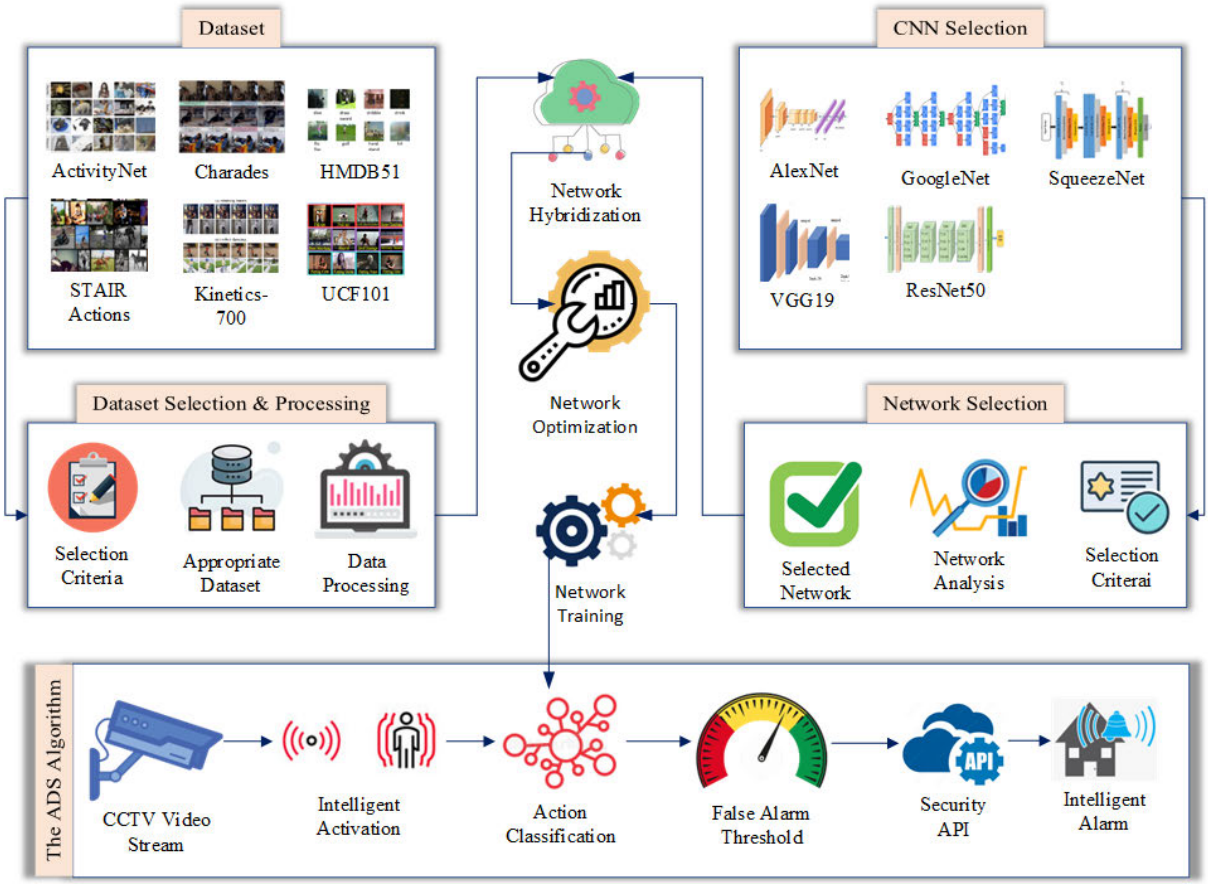


FIGURE 1. The overview of the proposed methodology.

video datasets. However, it is dependent on the Microsoft Kinect camera. It also requires the Inertial Measurement Unit (IMU). These devices are expensive. On the other hand, our methodology is not dependent on any particular device. Even a low-budget webcam is enough to ensure advanced front-door security using our system.

III. METHODOLOGY

The proposed ADS algorithm uses a GoogleNet-BiLSTM hybrid network as the classifier. This hybrid network requires a video dataset. The video dataset selection criteria, dataset processing, network architecture, the working principle of the network with necessary mathematical interpretation, and the FDS algorithm have been described in this section. The overview of the proposed methodology has been illustrated in figure 1.

A. DATASET

The quality and relevance of the dataset play a significant role in the overall performance of the machine learning model, including the proposed network [27]. That is why selecting an appropriate dataset based on the criteria determined by the goals of an experiment is an essential step in CNN-based research. This sub-section presents our process of selecting the most appropriate dataset and methods of processing it.

TABLE 1. Description of the incidents and class names.

Serial	Incident	Class
1	Punching the door	Punch
2	Kicking the door	Kick
3	Hitting the door	Hit
4	Shooting at the door	Shoot

1) DATASET SELECTION CRITERIA

Predicting security breaches in real-time video streams is a broad field of research. We have narrowed it down to a front-door security breach. Various activities are subject to CCTV footage analysis to predict security threats. However, activities that impose a threat at the front door are limited. The possible incidents at the front doors and their class names are listed in the table 1.

The target class names are the primary dataset selection criteria. The target datasets are the Human Activity Recognition (HAR) related dataset, which has a rich collection of punching, kicking, hitting, and shooting samples. Based on social observation, the activities presented in table 1 have been selected. Anyone standing at the front door holding a gun is a security threat. It has been selected from the context

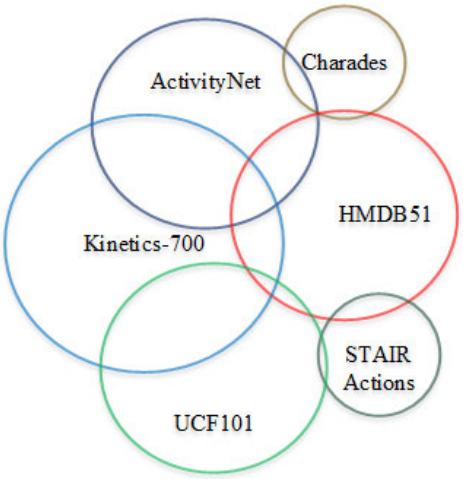


FIGURE 2. Class overlapping among experimenting datasets.

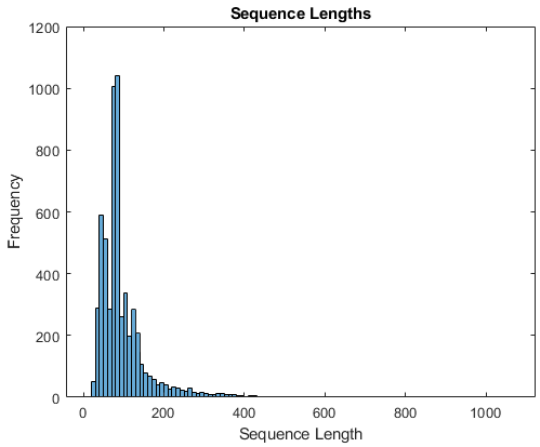


FIGURE 3. Video clip sequence lengths analysis.

of increasing gun violence. Another typical incident at the front door is hitting the door knob to break it to enter the house. Usually, burglars target empty houses and try to gain access by breaking the door. From this context, we have included the hitting on the door in the incident list. Anyone furious with the intention to physically hit someone usually punches or kicks the front door to express his anger. It is a common human nature. We have selected punching and kicking as target incidents considering these social phenomena.

This paper explores the potential of the available HAR dataset instead of creating one for the experiment. There are mainly three criteria we analyzed while selecting the dataset. These criteria have been set from the context of ease of implementation of the proposed FDS. That is why the availability of activities listed in table 1 is a critical selection criterion. The secondary criterion is the similarity of the video of the selected activities with the front-door environment. And the third criterion is the feature richness of the available videos.

2) DATASET SELECTION

The performance of Convolutional Neural Networks (CNNs) depends on both network architecture and training datasets [2]. We studied and analyzed six video datasets related to Human Activity Recognition (HAR) and listed them in table 2 [24].

We explored the datasets listed in table 2. It has been observed that there are overlapping classes among these datasets, which have been illustrated in figure 2. The ActivityNet and Kinectics-700 share the highest number of overlapping classes. That means these two datasets have a strong correlation with the HMDB51 dataset. The STAIR Actions dataset also shares some common classes. The UCF101 and the Charades dataset have the least common classes. As a result, we conclude that these two datasets maintain a weak correlation with the HMDB51 dataset.

Among these datasets, the HMDB51 dataset contains the target classes mentioned in table 1 along with 47 other classes. This dataset has 103 clips of gun shooting, 126 clips of punching, 130 clips of kicking, and 127 clips of hitting.

3) HAR VS. FRONT DOOR ACTIVITIES

The HMDB51 datasets have been selected as the primary dataset for this paper. It is a Human Activity Recognition (HAR) dataset. The proposed FDS algorithm focuses on front-door security only. The activities which are considered threats at the front door are subsets of the HMDB51 dataset. This dataset contains videos of punching (p), kicking (k), hitting (h), and holding guns (g), along with 47 other classes. A model trained using the HMDB51 dataset is suitable for classifying the activities listed in table 1. The equation 1 defines the relation between the set of selected activities and the HMDB51 dataset.

$$A = \{x | x \in H, (x \cap H) = \{p, k, h, g\}\} \tag{1}$$

Here in equation 1, A is the set of target activities, and H is the set of activities available in the HMDB51 dataset.

4) DATA PROCESSING

The video clips of the HMDB51 do not require any low-level filtering or improvement [34]. That means the dataset is ready to train the LSTM network. However, we have observed that some of the video clips are very lengthy. We analyzed the lengths using a histogram illustrated in figure 3.

The histogram shows that there are few lengthy video clips. The lengthier the videos, the longer the network takes to learn. Training machine learning model using image datasets is time-consuming. It takes even longer for a video dataset. Longer period of training refers to the occupation of computing resources for a longer period of time. At the same time, retraining the model for newer datasets becomes a serious issue as well. This experiment explored the opportunity to reduce the training time by limiting the length of the video. Training the proposed GoogleNet-BiLSTM hybrid network with the original HMDB51 dataset takes three hundred and fifty-eight minutes. The learning curve with the original HMDB51 dataset is presented in figure 4. It has been

TABLE 2. Human Activity Recognition (HAR) Datasets.

Dataset	Samples	Classes	Activities
ActivityNet [28]	21,313	200	Daily, social and household activities including sports and exercises
Charades [29]	66,493	157	Regular indoor activities such as filling cups, folding towels, etc.
HMDB51 [30]	5,100	51	Body movement, facial actions and interaction with objects
Kinetics-700 [31]	530,336	700	Single-person, and person-to-person interactions.
STAIR Actions [32]	109,478	100	Regular indoor activities in the home, office, washroom, kitchen, object manipulation, etc.
UCF101 [33]	13,320	101	Human-object interaction, body motion without object interaction, and using instruments.

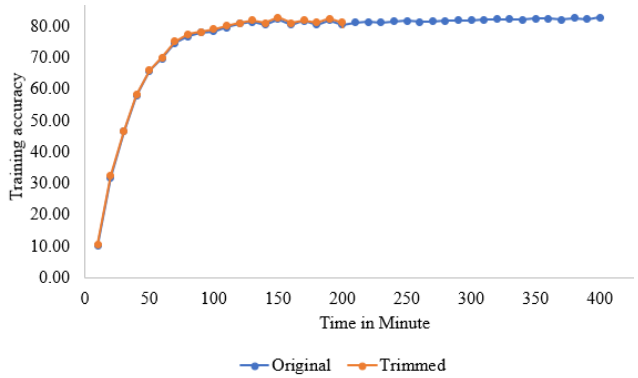


FIGURE 4. Learning curve with original HMDB51 dataset and trimmed dataset.

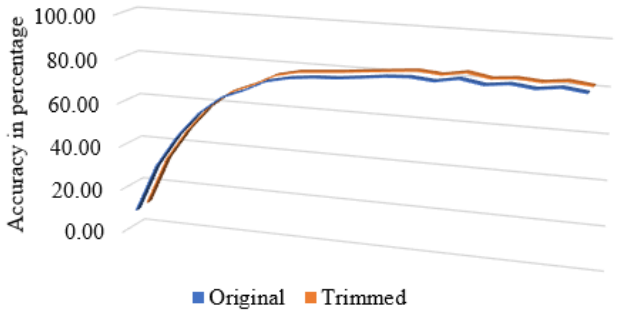


FIGURE 5. Performance comparison for the original and trimmed dataset.

observed from the learning curve that the accuracy of the proposed hybrid model does not have any significant change after two hundred minutes. However, the training process requires more iterations because of the video length. It is observable that the video length no longer positively impacts the model's performance. The average training accuracy is 73.24%, which experience ±0.06% deviation for epochs after two hundred minutes. From this observation, we have concluded that limiting the video length would make the training process faster with negligible effect on the overall performance. The effect of trimming video length has been presented on figure 4 as well.

It validates the claim of the insignificance of the video length trimming on the impact of the learning process of the proposed GoogleNet-BiLSTM hybrid network. Both learning curves show similar characteristics. Roughly after one hundred minutes of training, the network learning progress is almost flattened. And the average classification accuracy of the proposed network trained with a trimmed video length dataset is 73.18%. However, it takes 49.22% less time to complete the training process. Considering the necessity of frequently retraining the network for the updated dataset, the trimmed dataset has been used in this experiment.

The validation accuracy of the proposed network trained with the original and trimmed HMDB51 dataset has been illustrated in figure 5. It shows the insignificant variations between the performances. With the cost of ±0.06% impact on the validation accuracy, 50.78% training time is saved. It is an accuracy-time trade-off. This paper emphasizes reducing the training time by using the trimmed video length dataset. That is why videos of more than 300 seconds have been removed from the dataset.

B. NETWORK ARCHITECTURE

The proposed network combines a pretrained Convolutional Neural Network (CNN), GoogleNet, and a Long Short Term Memory (LSTM) network.

1) SEQUENCE FOLDING

The inputs to the network are video feeds which are sequences of image frames that maintain specific temporal distribution [35]. The features of the video frames need to be extracted to train the network. However, the feature extraction delay of the network and the temporal distribution of the video frames are not the same. As a result, video feature extraction from the stream of the video frame directly is not a realistic approach. A study by A. George & A. Ravindran shows that the latency for machine vision can be controlled through approximate computing [36]. However, we've taken a much easier solution to reduce the computational cost. This problem has been handled using sequence folding defined by equation 2.

$$\sum_{i=1}^{N} I(x_i, y_i) = \sum_{t=1}^{T} f_r((x_t, y_t), t) \tag{2}$$

Here $f_r((x_t, y_t), t)$ is a time-dependent frame and $I(x_i, y_i)$ is a time-independent image. The video frame sequence is converted into separate images using a folding sequence layer. These images are used to extract the features.

2) FEATURE EXTRACTOR NETWORK

The folded sequences contain video features. We used GoogleNet, a pretrained CNN, to extract the features [37]. We experimented with five popular pretrained networks listed in table 3.

TABLE 3. The pretrained CNNs experimented with in this paper.

Network	Depth	Size (MB)	Parameters (millions)	Input Size
AlexNet [38]	8	227	61.0	227x227
GoogleNet [8]	22	27	7.0	224x224
ResNet50 [39]	50	96	25.6	224x224
VGG19 [40]	19	535	144.0	224x224
SqueezeNet [41]	18	5.2	1.24	227x227

The performance of these pretrained networks has been demonstrated in the Result and Performance section. Despite the acceptable performance, we excluded VGG-19 and AlexNet for their size. The SqueezeNet is very lightweight and has significantly fewer learnable parameters. However, it lowers the accuracy of the overall network. GoogleNet outperforms Squeezenet and ResNet50. The difference in the final classification accuracy for VGG-19, AlexNet, and GoogleNet is almost identical. That is why, considering everything, we used GoogleNet in the algorithm 1 to extract the feature vectors.

Algorithm 1 Frames to Feature Vectors

Input: GoogleNet, G_N; Dataset, D_s
Output: Feature Vector Sequence, s;
Start
$I_s \leftarrow Size(Layers(1, G_N))$
$L \leftarrow pooling(N = 5, K = 7 \times 7, S = 1)$
$F \leftarrow num(readfiles(Dataset))$
for $i \leftarrow 1 : F$ **do**
  $v \leftarrow readVideo(file(i))$
  $s(i, 1) \leftarrow activations(G_N, v, L)$
end for
$save(s, FeatureVector)$
end

The features are saved as feature vectors. Once the features are extracted, the folded sequences are unfolded by inverting the equation 2. The video frames and corresponding feature vectors are 2D-spatial signals. They are flattened using a flatten layer before sending to the LSTM network.

3) GoogleNet-LSTM HYBRID NETWORK DESIGN

The proposed Googlenet-BiLSTM hybrid network has been developed by combining part of the GoogleNet and a BiLSTM network. The GoogleNet has been used as a feature extractor, and the BiLSTM network is responsible for the classification. The GoogleNet is a 22-layer deep convolutional neural network. The 19^{th} layer of the GoogleNet is an average pooling layer with a size of 7×7. This layer passes the extracted features to the subsequent Fully Connected (FC) layer [42]. However, the proposed hybrid network removes the layers after the pooling layer and passes the features to the BiLSTM network. We used a BiLSTM network with 2,000 hidden nodes illustrated in figure 6. The classification layer is connected to the BiLSTM layer through a fully connected layer [43]. In this network, we used the Softmax [44] activation function for the classification layer.

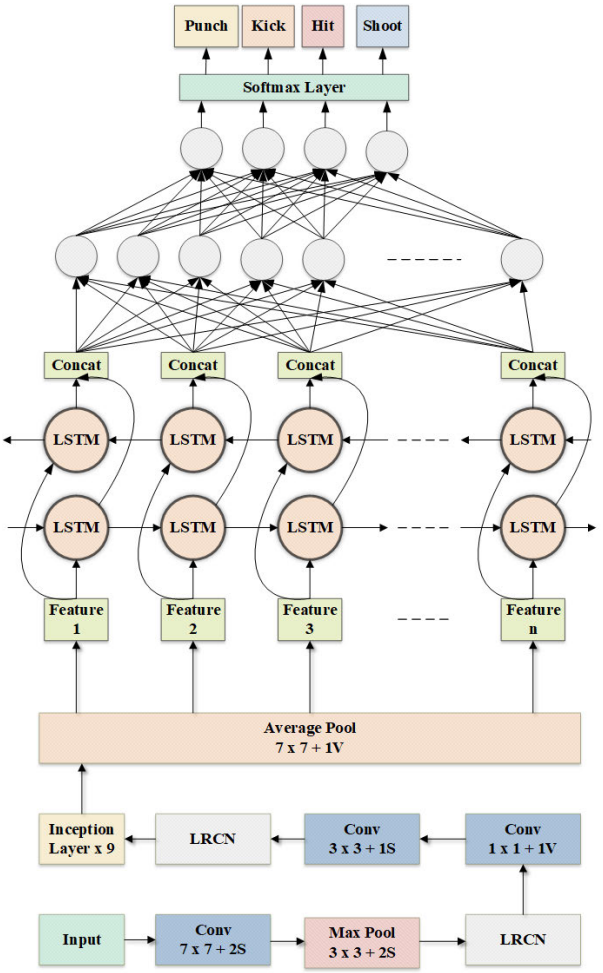


FIGURE 6. The GoogleNet-BiLSTM Hybrid Network.

The classification layer classifies the input video into one of the four classes - Punch, Kick, Hit, or Shoot.

4) TRAINING THE LSTM

We split the dataset into 80:10:10 for training, testing, and validation, respectively. We used the mini-batch method with a size 16 for each batch. The videos of each batch are internally shuffled in every iteration. Along with shuffling, we used $k-$fold cross-validation at $k = 4$.

We experimented with three optimization algorithms - Adaptive Gradient Algorithm (AdaGrad) [45], Root Mean Squared Propagation (RMSProp) [46], and Adaptive Moment Estimation (ADAM) [47] defined in equation 3, 4, and 5, respectively.

$$\omega_i^{(t+1)} = \omega_i^t - \frac{\eta}{\sqrt{\sum_{\tau=1}^{t} g_{\tau,i}^2}} g_{t,i} \tag{3}$$

$$\omega_i^{(t+1)} = \omega_i^t - \frac{\eta}{\sqrt{(v_t) + \epsilon}} \Delta_t \tag{4}$$

$$\omega_i^{(t+1)} = \omega_i^t - m_t\left(\frac{\alpha}{\sqrt{v_t} + \epsilon}\right) \tag{5}$$

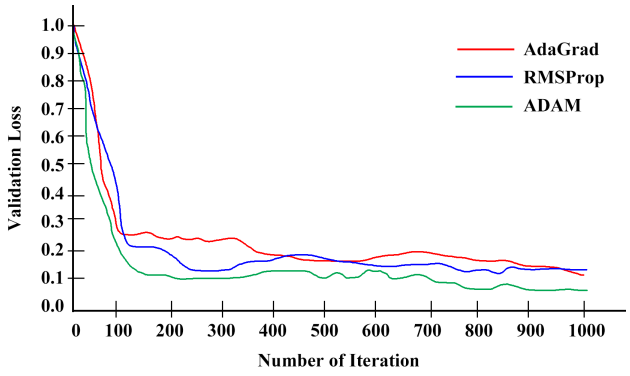


FIGURE 7. The optimization algorithm selection.

We analyzed the validation loss curve with respect to the number of iterations illustrated in figure 7. It shows that the ADAM performs better than AdaGrad and RMSProp algorithms.

The validation loss rapidly reduces up to 120 iterations for each optimization algorithm. After that, the loss reduces gradually to 1000 iterations. The validation loss is the lowest for the ADAM. That is why it has been used in our network. The training progress is illustrated in figure 8.

The validation accuracy during the training progress increases rapidly till 120^{th} iterations. After that, the rate of increment of validation accuracy reduces. However, it does not stop, and it keeps increasing till 1000^{th} iterations. Similar but inverse nature is observed for the validation loss. It rapidly reduces till 120^{th} iteration. After that, the rate of change gradually reduces. The training process completes after 1000 iterations with 81.16% validation accuracy. The training process takes 197 minutes and 15 seconds with a 0.0001 learning rate.

5) FRONT-DOOR SECURITY (FDS) ALGORITHM

The FDS written in algorithm 2 uses the trained BiLSTM network to classify the four actions mentioned in table 1. This algorithm reads the real-time video stream. When the camera is active, it reads the frames. If a significant difference exists between two successive frames, the algorithm calls the BiLSTM network and starts passing the frames. The BiLSTM network predicts the label of the action (p) in the video with a confidence score (s). Based on the predicted label and confidence score, the algorithm uses the security application API to alert the user. If there is no significant difference between two successive frames, the algorithm takes no action.

The thresholds of the proposed FDS algorithm play significant roles in the overall performance. We experimented with threshold values between 0 to 1 with 0.1 increments. It has been observed that a threshold below 0.6 impacts the algorithm's performance for every class. In this range, the average accuracy is 34.10% which is not feasible for a security system. However, there is a significant improvement within the threshold range of 0.6 to 0.95. The analysis in between this range has been presented in figure 9. It is evident that the thresholds 0.70, 0.72, 0.65, and 0.85 generates the best result for punch, kick, hit, and gun classes, respectively. The threshold has been selected based on the average performance of the proposed algorithm on all experimenting datasets listed in table 2.

Algorithm 2 The FDS Algorithm

Input: CCTV Video Stream, v_s;
Output: Alert, a;
Start
$i \leftarrow 0$
$F[i] \leftarrow read(v_s)$
while $v_s = True$ **do**
  $i \leftarrow i + 1$
  $F[i] \leftarrow read(v_s)$
  $c \leftarrow compare(F[i-1], F[i])$
  if $c \geq 0.5$ **then**
    $[p, s] \leftarrow LSTM[F[i]]$
    if $p == Punch \ \& \ s \geq 0.70$ **then**
      $a \leftarrow DoorPunch$
    else if $p == Kick \ \& \ s \geq 0.72$ **then**
      $a \leftarrow DoorKick$
    else if $p == Hit \ \& \ s \geq 0.65$ **then**
      $a \leftarrow DoorHit$
    else if $p == Shoot \ \& \ s \geq 0.85$ **then**
      $a \leftarrow DoorShooting$
    end if
    $SecurityAPI(a)$
  else
    $NoAction$
  end if
end while
end

IV. RESULTS AND PERFORMANCE EVALUATION

The performance of the FDS algorithm depends on the accuracy of the proposed GoogleNet-BiLSTM hybrid network. As it is Deep Learning (DL) approach, we used state-of-the-art machine learning performance evaluation metrics [48] to assess the performance of the proposed network first. After that, we analyzed the performance among different models. This performance comparison demonstrates the superiority of the proposed methodology. After that, we performed another experiment with different datasets. The purpose of this third experiment is to analyze the robustness of the system. The proposed model has been trained with the HMDB51 dataset. However, we have used videos from all datasets mentioned in table 2 to investigate the performance of the proposed hybrid network.

1) OVERALL PERFORMANCE

This section covers the performance of the GoogleNet-BiLSTM hybrid network. The performance evaluation metrics we used are listed in table 4. The metrics, their

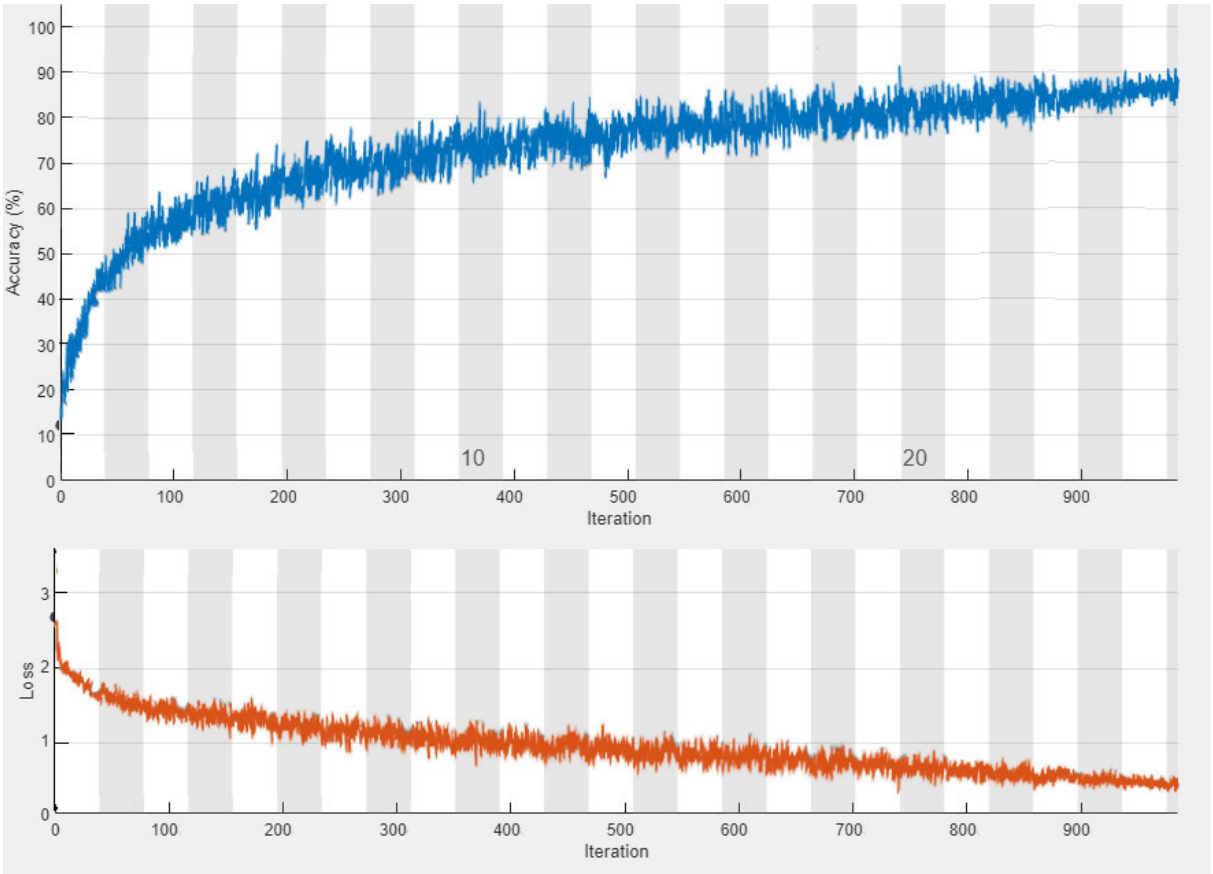


FIGURE 8. The training progress and related information.

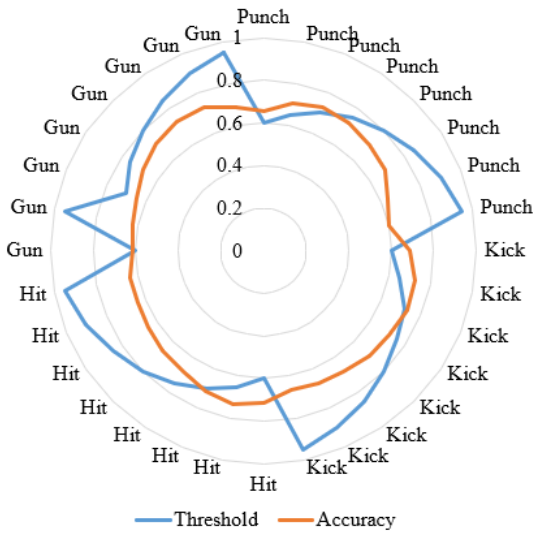


FIGURE 9. The threshold value analysis for Gun, Punch, Hit, Kick class with 0 to 1.0 range.

mathematical expression, and their roles in the evaluation are listed in the same table.

The metrics listed in table 4 requires the values of True Positive(*TP*), True Negative(*TN*), False Positive(*FN*), and

TABLE 4. Performance evaluation metrics.

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

False Positive(*FP*). We have generated a confusion matrix, illustrated in figure 10, to obtain these values.

Based on these values using the mathematical expression of the evaluation metrics, the performance of the proposed methodology has been listed in table 5. It shows the accuracy, sensitivity, FPR, and FNR separately for every class. The average values of these metrics are 73.18%, 68.25%, 69.78%, and 71.10%, respectively.

We also calculated the precision, recall, and F1-score of four different classes directly from the confusion matrix. The values of these evaluation metrics are listed in table 6.

The average precision is 0.7149. It proves the quality of the positive predictions made by our network. The ratio of the correctly identified positive class and the total number of

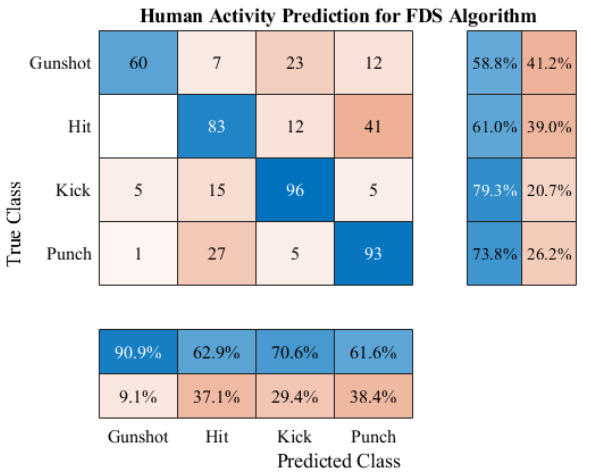


FIGURE 10. The confusion matrix for performance analysis.

TABLE 5. Performance of the classifier.

Activity	Accuracy	Sensitivity	Specificity	FPR	FNR
Gunshot	85.1%	68.20%	65.80%	34.2%	31.8%
Hit	65.0%	64.55%	63.19%	36.81%	35.45%
Kick	72.5%	74.04%	79.25%	20.75%	25.96%
Punch	70.1%	72.33%	76.10%	23.9%	27.67%

TABLE 6. Performance analysis using the confusion matrix.

Class	Precision	Recall	F1-score
Gunshot	0.9091	0.5882	0.7134
Hit	0.6288	0.6103	0.5119
Kick	0.7059	0.7934	0.7471
Punch	0.6159	0.7381	0.6064

positive samples is 0.6825. And the harmonic mean is 0.6447. The statistical analysis of these numeric values proves that the performance of the proposed hybrid network is acceptable. A comparison study in the following section supports this claim.

2) PERFORMANCE COMPARISON AMONG DIFFERENT MODELS

We experimented with Multi-layer Perceptron (MLP) [49], Convolutional Neural Network(CNN) [50], Long Short Term Memory Sequence (LSTM) [51], and Bidirectional Long Short-Term Memory (BiLSTM) [52]. In this experiment, 25 video clips from each class were randomly selected. After that, they are further categorized into 30 seconds and 60 seconds categories. The experimenting video sequences were curved to 30 and 60 seconds for each model to create a fairground. The accuracy, precision, recall, and F1-score for 30 and 60 seconds video clips are listed in table 7.

The numerical data in table 7 shows that the performance of the proposed network is slightly better than BiLSTM. Because of using GoogleNet to extract the video features, the network's performance was 2.80%. It highlights the contribution of our methodology in improving the human action recognition accuracy from the CCTV video frame streams. That means the proposed model, a combination of CNN and BiLSTM, performs better than other models. It has been observed that the performances on shorter videos are better than those on longer videos. However, in either of the cases, the proposed model performs better.

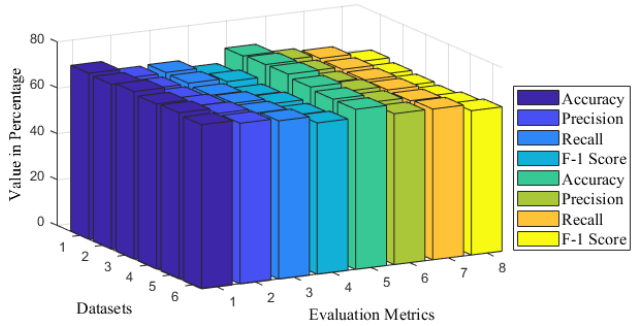


FIGURE 11. The demonstration of insignificant performance variations of the proposed system for different datasets.

3) PERFORMANCE COMPARISON AMONG DIFFERENT DATASETS

We studied seven different datasets before developing the methodology. The proposed network has been trained, tested, and evaluated using the HMDB51 dataset. To further evaluate the performance of the proposed network, we experimented with the network with the seven datasets listed in table 2. In this experiment, we choose similar classes only. After that, the video clips from each class were randomly selected. After random selection, the videos were trimmed into 30 seconds and 60 seconds clips. This experiment used 20 video clips from each class. The result of this experiment has been listed in table 8.

The experimenting network performs well on each dataset mentioned in table 8. In this experiment, we used video clips that belong to the same class. We experimented with four classes where shooting a gun is a class. The other datasets do not have this class. However, there are samples of holding objects similar to guns. We used these classes to compensate for the limitations of the experimenting datasets. The results obtained from this experiment show an average accuracy of 72.21%. That means there are insignificant accuracy differences for different datasets. That means the proposed network is robust, and there are no overfitting issues. It is not biased to any particular dataset. Figure 11 demonstrates the nominal differences among the same evaluation metric of different datasets. In this figure, the dataset is along the x-axis that maintains the sequence of table 8.

V. REAL-WORLD IMPLEMENTATION

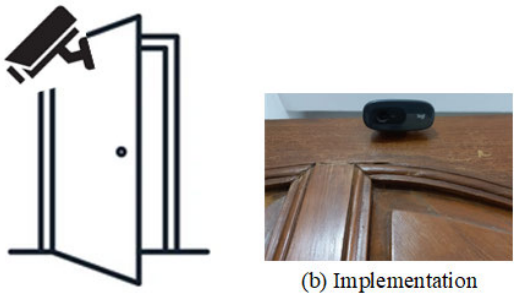
The performance of the proposed front door security system in the laboratory experiment is satisfactory. However, there are always some differences between laboratory and real-world scenarios. We have implemented the security system at the front door of an apartment, which is illustrated in figure 12. We used a Logitech C270 HD webcam. It has been mounted at the top of the door at 45^o viewing angle

TABLE 7. Performance comparison among different models.

Model Name	Frame Sequence							
	30 Seconds Clips				60 Seconds Clips			
	Accuracy	Precision	Recall	F-1 Score	Accuracy	Precision	Recall	F-1 Score
BiLSTM	71.13%	70.05%	67.84%	63.55%	70.22%	68.75%	**66.78%**	60.52%
CNN	65.11%	64.00%	63.47%	59.42%	63.15%	63.47%	60.73%	56.44%
MLP	69.04%	63.14%	62.95%	58.82%	65.10%	62.80%	61.14%	54.27%
LSTM	68.46%	65.82%	67.13%	63.11%	67.99%	61.72%	65.48%	60.16%
Proposed Model	**73.18%**	**71.49%**	**68.25%**	**64.47%**	**71.25%**	**70.44%**	63.70%	**62.70%**

TABLE 8. Performance comparison among different datasets.

Dataset	Frame Sequence							
	30 Seconds Clips				60 Seconds Clips			
	Accuracy	Precision	Recall	F-1 Score	Accuracy	Precision	Recall	F-1 Score
ActivityNet	72.39%	68.23%	68.50%	65.20%	71.48%	**67.45%**	**66.20%**	62.65%
Charades	71.88%	67.87%	68.14%	**66.71%**	72.05%	66.80%	65.73%	61.44%
HMDB51	73.05%	67.92%	67.23%	64.88%	**72.19%**	66.32%	65.44%	60.78%
Kinetics-700	71.94%	69.01%	68.04%	65.72%	70.88%	67.07%	65.39%	60.29%
STAIR Actions	**72.42%**	70.00%	68.11%	65.09%	69.90%	66.46%	64.99%	61.62%
UCF101	71.55%	**70.05%**	**69.02%**	66.00%	69.79%	65.77%	65.70%	**62.86%**



(a) Design (b) Implementation

FIGURE 12. The camera mounted front door.

with respect to the door. We set the camera at 30 Frame Per Second (FPS). We avoided wireless communication and set up the system using a USB cable to eliminate additional delay caused by the wireless network. The camera is connected to a laptop computer where the proposed FDS algorithm is installed.

1) OBSERVATION & ANALYSIS

A glimpse of the observation on the performance of the proposed FDS algorithm in the real world has been illustrated in figure 13. The performance was observed over a period of 600 seconds, where an actor randomly kicked, hit, punched, and drew a gun at the door. According to human observation, the FDS algorithm effectively identifies the target incidents. We have also analyzed the proposed system's performance in the real world numerically. In this analysis process, we recorded the video of each incident for 150 seconds. Then we observed the time duration of the FDS system correctly and incorrectly identified the incidents. Based on the ratio of the correct and incorrect identification time duration, a performance score has been generated. The mathematical equation to generate the performance score is

TABLE 9. Performance score based on observational analysis.

Serial	Class	Observation	Score
1	Punch	Punching on the door	0.95
2	Kick	Kicking on the door	0.91
3	Hit	Hitting on the door	0.91
4	Shoot	Gun violence	0.96

defined by equation 6.

$$S = \frac{C_i}{I_i} \tag{6}$$

The C_i and I_i of equation 6 refer to Correct Identification Duration and Incorrect Identification Duration, respectively. The performance score for the experimenting incidents according to the real-world observation has been listed in table 9. According to the performance score, the proposed front door security algorithm using GoogleNet-BiLSTM hybridization is an effective system to strengthen the security at the front door.

VI. LIMITATION AND FUTURE SCOPE

No computing system is immune to limitations. The proposed FDS is not an exception. Despite good performance in solving real-world problems, it has several limitations worth attention.

1) SUBJECT-CAMERA ANGLE SENSITIVITY

One of the limitations of the proposed methodology is subject-camera angle sensitivity. It has been observed that if the subject is too close to the camera, that means the angle is more comparable to 70 degrees; the accuracy of the kicking and punching classes increases. However, the accuracy of the remaining classes decreases. When the angle is between 40 degrees to 60 degrees, the system generates the best result. The performance keeps degrading when the angle

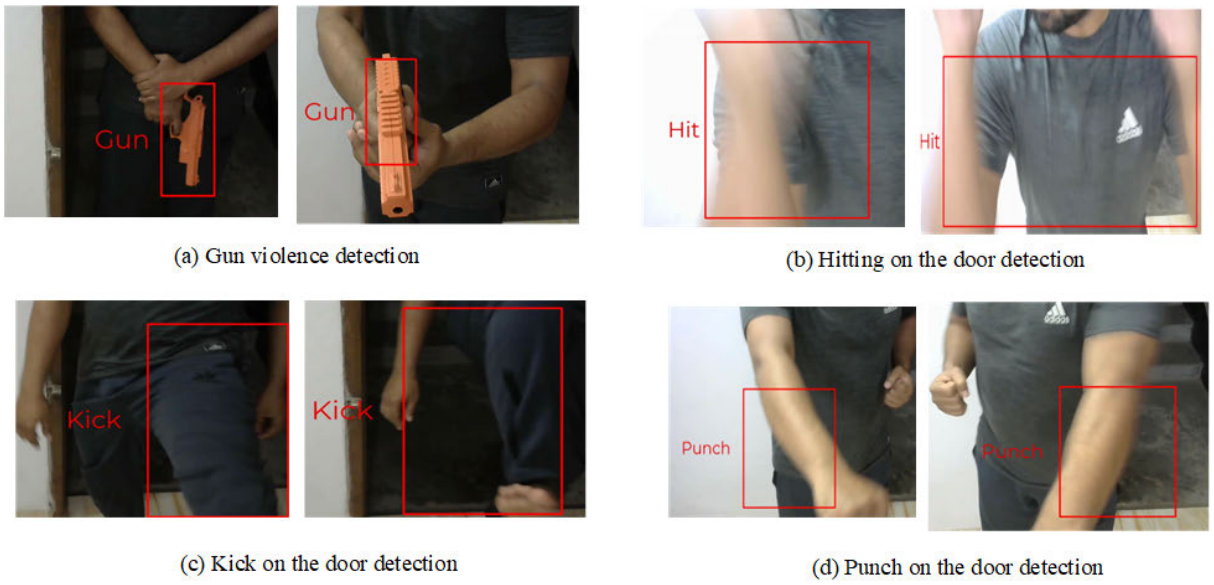


FIGURE 13. Detecting (a) Gun violence, (b) Hitting on the door, (c) Kicking on the door, and (d) Punching on the door using the FDS algorithm.

keeps increasing more than 60 degrees. The experimental results presented in this paper have been measured between 40 and 60 degrees. This limitation opens the opportunity to conduct further research and make the FDS more robust.

2) NIGHT-VISION EXPERIMENT

Security systems are for both day and night. The proposed methodology has been experimented with in daylight when the sun is up and in floodlights at night. Experimenting with night-vision mode is another challenge that this paper has not addressed. However, subsequent research is in progress to explore the performance in night-vision mode.

3) ACTIVITY LIMITATION

This paper deals with four security threats. There are many other potential risks at the front door which we could not include because of data availability. A custom dataset designed for the FDS algorithm would allow us to detect all of the common security threats at the front door. It paves the path to conduct more research, including new dataset creation and processing, discovering limitations and improving them for new classes, and handling computational complexities for additional classes.

The limitations of the proposed FDS algorithm pave the path to develop this system further. Instead of considering these as limitations, the researchers of this project consider them as opportunities to continue the research work and integrate more innovative features with it.

VII. CONCLUSION

Human Activity Recognition (HAR) has drawn attention from the research of wearable sensors and the computer vision scientific community. In this paper, we created a hybrid network combining state-of-the-art techniques found in current research trends. And our innovative approach is a potential solution to better front-door security. The advancement of research in HAR is eye-catching. However, its affordable application in front door security is unexplored. There are expensive, and large-scale AI surveillance services available that use HAR technology to strengthen the security of large premises. However, these services require expensive infrastructure. The FDS algorithm we presented in this paper does not require additional equipment. Integrating the CCTV camera video stream or a simple webcam is enough to recognize the security threats with 73.1% accuracy with an optimized threshold to reduce the false alarm rate. The real-world implementation and its experimental results show the adaptability of the FDS system in strengthening the security at the front door. However, there are some limitations of the FDS algorithm alongside impressive performance. These limitations are further opportunities to improve the system's service quality and robustness to make intelligent front-door security affordable and available for everyone.

REFERENCES

[1] A. B. Malali and S. Gopalakrishnan, "Application of artificial intelligence and its powered technologies in the Indian banking and financial industry: An overview," *IOSR J. Hum. Social Sci.*, vol. 25, no. 4, pp. 55–60, 2020.
[2] N. Faruqui, M. A. Yousuf, M. Whaiduzzaman, A. K. M. 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," *Comput. Biol. Med.*, vol. 139, Dec. 2021, Art. no. 104961.
[3] W. Xu and F. Ouyang, "The application of AI technologies in STEM education: A systematic review from 2011 to 2021," *Int. J. STEM Educ.*, vol. 9, no. 1, pp. 1–20, Sep. 2022.
[4] N. C. Eli-Chukwu, "Applications of artificial intelligence in agriculture: A review," *Eng., Technol. Appl. Sci. Res.*, vol. 9, no. 4, pp. 4377–4383, 2019.
[5] S. Weibin, L. Yun, D. Yi, D. Yingguo, P. Mingbo, and X. Gang, "Three-real-time architecture of industrial automation based on edge computing," in *Proc. IEEE Int. Conf. Smart Internet Things (SmartIoT)*, Aug. 2019, pp. 372–377.

[6] R. Abduljabbar, H. Dia, S. Liyanage, and S. A. Bagloee, "Applications of artificial intelligence in transport: An overview," *Sustainability*, vol. 11, no. 1, p. 189, 2019.

[7] S. Laato, A. Farooq, H. Tenhunen, T. Pitkamaki, A. Hakkala, and A. Airola, "AI in cybersecurity education—A systematic literature review of studies on cybersecurity MOOCs," in *Proc. IEEE 20th Int. Conf. Adv. Learn. Technol. (ICALT)*, Jul. 2020, pp. 6–10.

[8] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[9] B. Tripp, "Approximating the architecture of visual cortex in a convolutional network," *Neural Comput.*, vol. 31, no. 8, pp. 1551–1591, Aug. 2019.

[10] K. Yousaf and T. Nawaz, "A deep learning-based approach for inappropriate content detection and classification of YouTube videos," *IEEE Access*, vol. 10, pp. 16283–16298, 2022.

[11] Y.-H. Hsieh and X.-P. Zeng, "Sentiment analysis: An ERNIE-BiLSTM approach to bullet screen comments," *Sensors*, vol. 22, no. 14, p. 5223, Jul. 2022.

[12] Y. Li and L. Wang, "Human activity recognition based on residual network and BiLSTM," *Sensors*, vol. 22, no. 2, p. 635, Jan. 2022.

[13] P. Chorghe, S. Pathak, and S. Joshii, "A review on intelligent surveillance system for crime detection," *Int. J. Recent Adv. Multidisciplinary Topics*, vol. 3, no. 4, pp. 25–28, 2022.

[14] A. Ali, W. Samara, D. Alhaddad, A. Ware, and O. A. Saraereh, "Human activity and motion pattern recognition within indoor environment using convolutional neural networks clustering and Naive Bayes classification algorithms," *Sensors*, vol. 22, no. 3, p. 1016, Jan. 2022.

[15] M. Fu, Q. Zhong, and J. Dong, "Sports action recognition based on deep learning and clustering extraction algorithm," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–9, Mar. 2022.

[16] F. Kulsoom, S. Narejo, Z. Mehmood, H. N. Chaudhry, A. Butt, and A. K. Bashir, "A review of machine learning-based human activity recognition for diverse applications," *Neural Comput. Appl.*, vol. 34, pp. 18289–18324, Aug. 2022.

[17] D. R. Beddiar, A. Hadid, B. Nini, and M. Sabokrou, "Vision-based human activity recognition: A survey," *Multimedia Tools Appl.*, vol. 79, no. 41, pp. 30509–30555, 2020.

[18] B. Sarp and T. Karalar, "Real time smart door system for home security," *Int. J. Sci. Res. Inf. Syst. Eng.*, vol. 1, no. 2, pp. 121–123, 2015.

[19] Andreas, C. R. Aldawira, H. W. Putra, N. Hanafiah, S. Surjarwo, and A. Wibisurya, "Door security system for home monitoring based on ESP32," *Proc. Comput. Sci.*, vol. 157, pp. 673–682, Jan. 2019.

[20] A. A. Zeman, J. B. Ritchie, S. Bracci, and H. O. D. Beeck, "Orthogonal representations of object shape and category in deep convolutional neural networks and human visual cortex," *Sci. Rep.*, vol. 10, no. 1, pp. 1–12, Feb. 2020.

[21] S. S. Satam and H. El-Ocla, "Home security system using wireless sensors network," *Wireless Pers. Commun.*, vol. 125, pp. 1185–1201, Mar. 2022.

[22] P. Banerjee, P. Datta, S. Pal, S. Chakraborty, A. Roy, S. Poddar, S. Dhali, and A. Ghosh, "Home security system using Raspberry Pi," in *Advanced Energy and Control Systems*. Springer, 2022, pp. 167–176.

[23] J. Tao, H. Wu, S. Deng, and Z. Qi, "Overview of intelligent home security and early warning system based on Internet of Things technology," *Int. Core J. Eng.*, vol. 8, no. 5, pp. 727–732, 2022.

[24] Y. Kong and Y. Fu, "Human action recognition and prediction: A survey," *Int. J. Comput. Vis.*, vol. 130, no. 5, pp. 1366–1401, May 2022.

[25] V. Mazzia, S. Angarano, F. Salvetti, F. Angelini, and M. Chiaberge, "Action transformer: A self-attention model for short-time pose-based human action recognition," *Pattern Recognit.*, vol. 124, Apr. 2022, Art. no. 108487.

[26] W. Qi, N. Wang, H. Su, and A. Aliverti, "DCNN based human activity recognition framework with depth vision guiding," *Neurocomputing*, vol. 486, pp. 261–271, May 2022.

[27] A. Jain, H. Patel, L. Nagalapatti, N. Gupta, S. Mehta, S. Guttula, S. Mujumdar, S. Afzal, R. S. Mittal, and V. Munigala, "Overview and importance of data quality for machine learning tasks," in *Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2020, pp. 3561–3562.

[28] F. C. Heilbron, V. Escorcia, B. Ghanem, and J. C. Niebles, "ActivityNet: A large-scale video benchmark for human activity understanding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 961–970.

[29] G. A. Sigurdsson, G. Varol, X. Wang, A. Farhadi, I. Laptev, and A. Gupta, "Hollywood in homes: Crowdsourcing data collection for activity understanding," in *Proc. Eur. Conf. Comput. Vis.* Springer, 2016, pp. 510–526.

[30] H. Kuehne, H. Jhuang, E. Garrote, T. Poggio, and T. Serre, "HMDB: A large video database for human motion recognition," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 2556–2563.

[31] J. Carreira and A. Zisserman, "Quo vadis, action recognition? A new model and the kinetics dataset," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6299–6308.

[32] Y. Yoshikawa, J. Lin, and A. Takeuchi, "STAIR actions: A video dataset of everyday home actions," 2018, *arXiv:1804.04326*.

[33] K. Soomro, A. R. Zamir, and M. Shah, "UCF101: A dataset of 101 human actions classes from videos in the wild," 2012, *arXiv:1212.0402*.

[34] V. Sharma, M. Gupta, A. K. Pandey, D. Mishra, and A. Kumar, "A review of deep learning-based human activity recognition on benchmark video datasets," *Appl. Artif. Intell.*, vol. 36, no. 1, Dec. 2022, Art. no. 2093705.

[35] M. Mendieta, C. Neff, D. Lingerfelt, C. Beam, A. George, S. Rogers, A. Ravindran, and H. Tabkhi, "A novel application/infrastructure co-design approach for real-time edge video analytics," in *Proc. SoutheastCon*, Apr. 2019, pp. 1–7.

[36] A. George and A. Ravindran, "Latency control for distributed machine vision at the edge through approximate computing," in *Proc. Int. Conf. Edge Comput.* Springer, 2019, pp. 16–30.

[37] R. K. Bhogal and V. Devendran, "Human activity recognition using LSTM with feature extraction through CNN," in *Smart Trends in Computing and Communications*. Springer, 2023, pp. 245–255.

[38] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, Jun. 2012.

[39] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[40] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[41] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size," 2016, *arXiv:1602.07360*.

[42] K. Teymournezhad, H. Azgomi, and A. Asghari, "Detection of counterfeit banknotes by security components based on image processing and GoogleNet deep learning network," *Signal, Image Video Process.*, vol. 16, pp. 1505–1513, Jan. 2022.

[43] T. N. Sainath, O. Vinyals, A. Senior, and H. Sak, "Convolutional, long short-term memory, fully connected deep neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2015, pp. 4580–4584.

[44] R. A. Dunne and N. A. Campbell, "On the pairing of the softmax activation and cross-entropy penalty functions and the derivation of the softmax activation function," in *Proc. 8th Aust. Conf. Neural Netw.*, vol. 181. Melbourne, VIC, Australia: Citeseer, 1997, p. 185.

[45] A. Lydia and S. Francis, "Adagrad—An optimizer for stochastic gradient descent," *Int. J. Inf. Comput. Sci.*, vol. 6, no. 5, pp. 566–568, 2019.

[46] S. K. Turitsyn, T. Schafer, and V. K. Mezentsev, "Generalized root-mean-square momentum method to describe chirped return-to-zero signal propagation in dispersion-managed fiber links," *IEEE Photon. Technol. Lett.*, vol. 11, no. 2, pp. 203–205, Feb. 1999.

[47] W. K. Newey, "Adaptive estimation of regression models via moment restrictions," *J. Econometrics*, vol. 38, no. 3, pp. 301–339, Jul. 1988.

[48] M. Hossin and N. Sulaiman, "A review on evaluation metrics for data classification evaluations," *Int. J. Data Mining Knowl. Manage. Process*, vol. 5, no. 2, pp. 1–11, Mar. 2015.

[49] M. Riedmiller and A. Lernen, "Multi layer perceptron," Mach. Learn. Lab. Special Lect., Univ. Freiburg, Breisgau, Germany, Tech. Rep., 2014, pp. 7–24.

[50] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *Proc. Int. Conf. Eng. Technol. (ICET)*, Aug. 2017, pp. 1–6.

[51] J. Brownlee, *Long Short-Term Memory Networks With Python: Develop Sequence Prediction Models With Deep Learning*. Vermont, VIC, Australia: Machine Learning Mastery, 2017.

[52] Y. Bin, Y. Yang, F. Shen, X. Xu, and H. T. Shen, "Bidirectional long-short term memory for video description," in *Proc. 24th ACM Int. Conf. Multimedia*, Oct. 2016, pp. 436–440.

LUIZ PAULO OLIVEIRA PAULA was born in Mogi das Cruzes, São Paulo, Brazil, in 1991. He received the bachelor's degree in computer science from Centro Universitario UniBTA (formerly know as the Brazilian Institute of Advanced Technology—IBTA).

He has been leading data science and analytics projects to deliver actionable insights and data-driven decisions using AI and machine learning algorithms focused on performance growth and process improvement for clients, such as Kimberly-Clark, Verisign, Sanofi, and MSD/Merck. In his lectures as a Professor, he had the opportunity to give in-company classes for Deloitte, Petrobras, Chubb, and GDM, among others. His research interests include user behavior and human activity recognition using computer vision and deep learning hybrid networks.

Mr. Paula is a member of the Association of Computing Machinery and the Association for the Advancement of Artificial Intelligence.

NURUZZAMAN FARUQUI received the B.Sc. degree in electrical and electronics engineering from North South University and the master's degree in information technology from the Institute of Information Technology (IIT), Jahangirnagar University (JU), Bangladesh, in 2018, with a 4/4 CGPA.

He is currently working as a Senior Lecturer with the Department of Software Engineering (SWE), Daffodil International University, Bangladesh. He is a Research Coordinator with the Department of Software Engineering. He is also a YouTuber and the author. He is globally recognized for his educational video content on neural networks using MATLAB. He has authored three books. His research interests include artificial intelligence, machine learning, deep learning, cloud computing, and image processing.

Mr. Faruqui is a member of The Institution of Engineers, Bangladesh (IEB) and the Bangladesh Society for Private University Academics (BSPUA).

IMRAN MAHMUD received the master's degree in software engineering from the University of Hertfordshire, U.K., in 2008, and the Ph.D. degree in technology management from Universiti Sains Malaysia, in 2017.

He was a Senior Lecturer at the Graduate School of Business, Universiti Sains Malaysia. He was a Visiting Lecturer at the Institute Technology, Bandung, Indonesia, and the Hong Kong Management Association Hong Kong. He is currently working as the Head and an Associate Professor with the Department of Software Engineering, Daffodil International University, Bangladesh. He is also working as a Visiting Professor with the Graduate School of Business, Universiti Sains Malaysia.

Dr. Mahmud achieved several awards, including the Hall of Fame and the Presitigious Publication Award from Universiti Sains Malaysia, the Young Researcher Award from Kasetsart University, Thailand, and the Young Scientist Award in Technology Management from the Venus International Foundation, India. At the moment, he is working with technostress, games addiction, and online learning continuance intention.

MD. WHAIDUZZAMAN (Senior Member, IEEE) received the bachelor's degree in electronics and computer science, the M.Sc. degree in telecommunication and computer network engineering, London, U.K., and the Ph.D. degree from the University of Malaya, Malaysia.

He works as a Professor with the Institute of Information Technology (IIT), Jahangirnagar University. He is currently working on ARC-Funded Projects with the Queensland University of Technology, Australia. His research interests include mobile cloud computing, vehicular cloud computing, fog computing, the IoT, and microservices.

Dr. Whaiduzzaman received the *Journal of Network and Computer Applications* (Elsevier) Best Paper Award in Paris, France.

ERIC CHARLES HAWKINSON is a learning futurist, tinkering and designing technologies that may better inform the future of teaching and learning. His projects have included augmented tourism rallies, AR community art exhibitions, mixed reality escape rooms, and other experiments in immersive technology. He is currently working as a Professor of learning technology with the Faculty of Global Engagement. Kyoto University of Foreign Studies, Japan, and an Adjunct Professor with the Faculty of Information and Communication Studies, University of the Philippines Open University.

SANDEEP TRIVEDI (Senior Member, IEEE) received the bachelor's degree in electronic and communication engineering from Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal, India, in 2005.

He is currently working as a Senior SAP Analytics Consultant with Deloitte Consulting LLP, U.S. He specializes in SAP analytics reporting using machine learning algorithms. His SAP certifications include multiple SAP technologies, and have participated in various implementation projects in India, Qatar, United Arab Emirates, Malaysia, Singapore, and the USA. His research interests include artificial intelligence, machine learning, and deep learning.