



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

# IJRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 4, April 2024

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

Impact Factor: 8.423

 9940 572 462

 6381 907 438

 [ijirset@gmail.com](mailto:ijirset@gmail.com)

 [www.ijirset.com](http://www.ijirset.com)

# Smart Door Security

**Sugumaran. V. R, P. Praveen, A. Tamilmani, R. Vengadesh**

Assistant Professor, E.G.S. Pillay Engineering College, Nagapattinam, India

UG Student, Department of Computer Science and Engineering, E.G.S. Pillay Engineering College, Nagapattinam, India

**ABSTRACT:** The paper introduces an innovative automatic Front Door Security (FDS) algorithm that utilizes Human Activity Recognition (HAR) to identify four distinct security threats at the front door in real-time video footage, achieving an accuracy rate of 73.18%. The algorithm employs a novel combination of the GoogleNet-BiLSTM hybrid network to recognize these activities, enabling it to promptly alert against attempts to breach the door through kicking, punching, or hitting. Additionally, the FDS algorithm proves effective in detecting gun violence at the front door, further enhancing security measures. Demonstrating promising potential, this HAR-based FDS algorithm ensures improved safety with a precision of 71.49%, recall of 68.2%, and an F1-score of 0.65.

## I. INTRODUCTION

The utilization of Artificial Intelligence (AI) has revolutionized various facets of our lives, spanning across sectors such as banking, healthcare, education, agriculture, industrial automation, transportation, and more. These sectors leverage AI technology to streamline processes, minimize human involvement, and automate services, thereby enhancing efficiency. While the application of AI in cybersecurity has garnered significant attention for its effectiveness, its potential in physical security remains relatively unexplored. This paper delves into the efficacy of employing AI to reinforce front-door security through an innovative algorithm named FDS. The focus lies on automating front-door security systems to reduce the need for continuous human monitoring. The proposed algorithm emulates human-like intelligence to detect instances of gun violence and attempts to breach the front door through physical force, promptly alerting residents akin to a vigilant security guard.

In this research, we developed a hybrid network that integrates BiLSTM and GoogleNet to give a front-door video surveillance system human-like intelligence. The Google researchers built GoogleNet, a 22-layer deep Convolutional Neural Network (CNN) that can extract picture features in a manner similar to that of the human visual cortex. Additionally, the Bidirectional Long Sort Term Memory (BiLSTM) can pick up on these traits to categorize films on YouTube, detect emotional states in people, and identify actions in people.

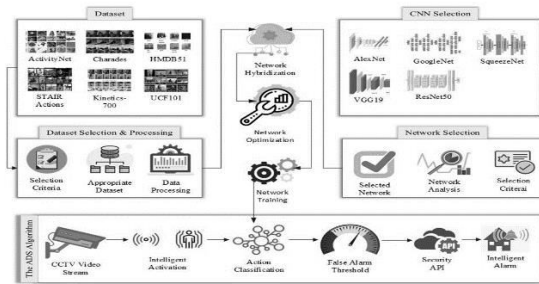
We found this to be an easy way to build an automated and intelligent entrance door security system. However, the reality is different and full of challenges and difficulties. In this paper, we present, analyze, and discuss the challenges we faced, the solutions we found, and the problems we overcame. Our discoveries, frontiers, and solutions deserve to be recognized as novel contributions in paving the way to ensuring automated and intelligent security systems are affordable for everyone. Avoiding false alarms is one of the challenges of intelligent surveillance systems. The higher the accuracy of human activity detection, the lower the chance of false alarms. However, in addition to accuracy, there are also other context-specific factors that determine the false alarm rate. The methodology described in this article detects human activity from live video feeds with acceptable accuracy and overcomes the problem of false positives. The main contributions of this paper are:

- Innovative Hybrid Network Design for HAR-Based Security System Development. achieved an average accuracy of 73.18%.
- Development of a new algorithm using CNN and its combination with BiLSTM for intelligent surveillance.
- A cost-effective solution that makes automated, intelligent security available to everyone with no nominal upfront or subscription fees.

## II. LITERATURE REVIEW

We reviewed the latest relevant literature on intelligent security systems based on human activity recognition (HAR). This shows the enviable progress of HAR and its applications in various fields. The performance of HAR systems

seems to have attracted the most attention from researchers, and their applications in the security field remain open for research. This paper focuses on the application of his HAR in information monitoring to enhance doorstep security. And HAR technology is the main driver of this approach. Therefore, our literature review focuses on advances in HAR through machine learning and its applications.



### A. Intelligent Application for Front Door Security

B. Sarp et al. used a Raspberry Pi-based video surveillance system to secure the front door through his two functions: video feed and communication. The company's system allows users to remotely monitor activity in front of their door and communicate with someone at the front door. In addition, we connected the door via the mobile phone network and accessed the functionality in real time via the Internet. Although this approach effectively ensures front door security, it also has drawbacks. The disadvantage is that manual inspection is required. His proposed FDS system does not require human intervention to monitor front door security. This is a fully automatic system that detects human activity at the door and alerts homeowners if something suspicious happens.

The ESP32-based home surveillance system presented by R.C. Aldawira et al. demonstrates the application of IoT to ensure home security, including front door security. This system allows users to remotely monitor activities inside and outside the home and control door locks. It also has a motion sensor that detects movement and notifies the user. It is also equipped with a touch sensor that detects when someone touches the doorknob. Many of these features make your home safer. However, this system does not use human-like intelligence. Because it uses motion and touch sensors, it has a high false alarm rate and requires manual adjustment. Compared to this approach, his proposed FDS is more advanced as it uses CNN and detects activity similar to human visual cortex. IoT-based home security systems, edge computer-based security systems, and intelligent alert-based security systems are common approaches to improve home security. A literature review highlights research gaps in entrance security using convolutional neural networks. The proposed FDS algorithm aims to fill this gap and leverage CNN to ensure human safety at the doorstep.

### B. Computer Vision-Based HAR and Applications

Computer vision-based human activity recognition is a key technology in video analytics and its applications in intelligent surveillance, self-driving cars, video analytics, video search, and entertainment. The reviews presented in this article are consistent without any observations or methodologies. Machine learning-based approaches based on computer vision are suitable for front door security algorithms. v. Mazzia et al. developed a short-term pose-based human action recognition system. With 227,000 parameters he achieved an accuracy of 90.86%. Although the accuracy of this paper is impressive, the high computational cost makes it unsuitable for developing affordable security systems using these methodological operations.

A DCNN-based framework guided by depth vision, as demonstrated by Q. Wen et al., achieved a commendable accuracy of 93.89%. This approach addresses the significant challenges associated with data collection and labeling for training machines on video datasets. However, it relies on the Microsoft Kinect camera and requires the use of an Inertial Measurement Unit (IMU), both of which are costly devices. In contrast, our methodology is device-agnostic, requiring only a low-budget webcam to implement advanced front-door security effectively.





### III. METHODOLOGY

The ADS algorithm outlined in this paper employs a GoogleNet-BiLSTM hybrid network as its classifier. To facilitate the operation of this hybrid network, a suitable video dataset is required. This section elaborates on the criteria for selecting the video dataset, details the dataset processing procedures, describes the network architecture, elucidates the operating principles of the network with requisite mathematical explanations, and presents the FDS algorithm.

#### A. DATASET

The quality and relevance of the datasets play an important role in the overall performance of the machine learning model containing the proposed network [27]. Therefore, selecting an appropriate dataset based on the criteria established by the purpose of the experiment is an important step for CNN-based methods.

TABLE 1. Description of the incidents and class names.

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

#### 1) CRITERIA FOR DATASET SELECTION

The task of forecasting security breaches within real-time video streams encompasses a wide array of research avenues. We have specifically focused on front-door security breaches within this domain. While CCTV footage analysis may encompass a range of activities for predicting security threats, our attention is directed towards activities specifically pertaining to threats encountered at the front door.

The primary criterion for selecting the dataset is the classification of target class names. Specifically, we focus on the Human Activity Recognition (HAR) dataset, which contains a comprehensive assortment of samples related to punching, kicking, hitting, and shooting. Through social observation, we have chosen the activities outlined in Table 1. It is important to note that any individual standing at the front door while holding a gun is considered a security threat.

#### 2) DATASET SELECTION

The effectiveness of Convolutional Neural Networks (CNNs) is influenced by both the network architecture and the training datasets. We conducted an examination of six video datasets associated with Human Activity Recognition (HAR), which are detailed in Table 2.

Our investigation focused on the datasets enumerated in Table 2. We observed that there are instances of overlapping classes among these datasets, as depicted in Figure 2. Notably, ActivityNet and Kinetics-700 exhibit the highest number of overlapping classes, indicating a significant correlation with the HMDB51 dataset. The STAIR Actions dataset also shares some common classes. Conversely, the UCF101 and Charades datasets display fewer common classes, suggesting a weaker correlation with the HMDB51 dataset.

Instead of creating a new dataset for experimentation, this paper investigates the potential of existing HAR datasets. Three main criteria were analyzed during the dataset selection process, focusing on facilitating the implementation of the proposed FDS. The primary criterion is the availability of activities listed in table 1. The secondary criterion considers the similarity of the selected activities' video footage to the front-door environment. Lastly, the third criterion evaluates the richness of features present in the available videos.

#### 3) HAR VERSUS FRONT DOOR ACTIVITIES

The HMDB51 dataset serves as the primary dataset chosen for this study, focusing on Human Activity Recognition (HAR). The proposed FDS algorithm specifically targets front-door security activities. These activities, regarded as threats at the front door, form subsets of the HMDB51 dataset. Within this dataset, videos encompassing punching (p), kicking (k), hitting (h), and holding guns (g), among 47 other classes, are included. A model trained using the



HMDB51 dataset is well-suited for categorizing the activities outlined in Table 1. Equation 1 outlines the relationship between the selected activities and the HMDB51 dataset:

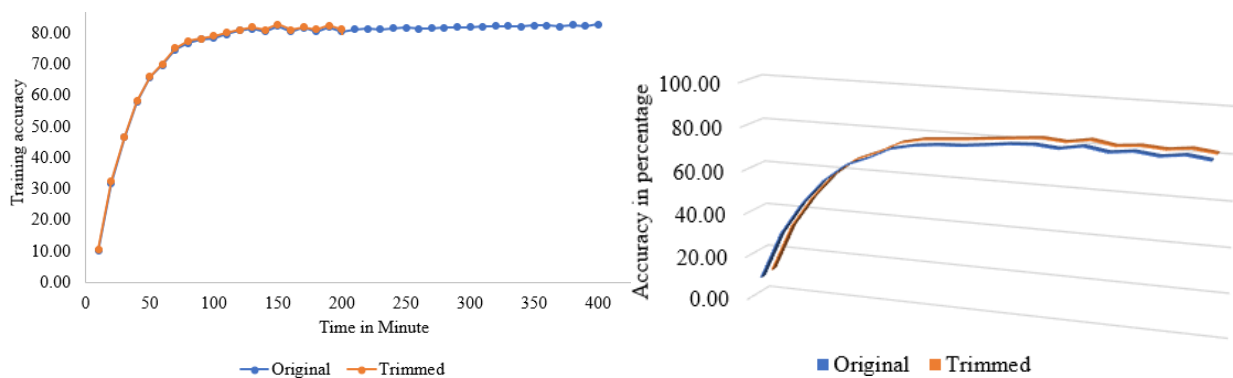
$$A = \{x \mid x \in H, (x \cap H) = \{p, k, h, g\}\}$$

In Equation 1, A represents the set of target activities, and H denotes the set of activities contained within the HMDB51 dataset.

#### 4) DATA PROCESSING

The video clips within the HMDB51 dataset do not require any extensive low-level filtering or enhancement, making the dataset suitable for training the LSTM network without additional preprocessing. However, upon analysis, we noticed that some video clips are excessively lengthy. A histogram, depicted in Figure 3, illustrates the distribution of video clip lengths.

The histogram reveals that a minority of video clips are notably lengthy. Lengthier videos translate to extended learning times for the network. Training machine learning models using image datasets is already time-consuming, and video datasets exacerbate this issue further. Lengthy training periods consume significant computing resources and pose challenges for retraining models with newer datasets. This experiment aimed to address these challenges by exploring methods to reduce training time, particularly by limiting the length of videos. Training the proposed GoogleNet-BiLSTM hybrid network using the original HMDB51 dataset required three hundred and fifty-eight minutes.



## B. NETWORK ARCHITECTURE

The proposed architecture integrates a pretrained Convolutional Neural Network (CNN), specifically GoogleNet, with a Long Short-Term Memory (LSTM) network.

### 1) SEQUENCE FOLDING

The network's inputs consist of video feeds, which are sequences of image frames maintaining specific temporal distribution. Extracting features from these video frames is essential for network training. However, there's a discrepancy between the feature extraction delay of the network and the temporal distribution of the video frames. Consequently, directly extracting video features from the frame stream is impractical. A study by A. George & A. Ravindran suggests that machine vision latency can be managed through approximate computing.

### 2) FEATURE EXTRACTOR NETWORK

The folded sequences encapsulate video features, for which we employed GoogleNet, a pretrained CNN, for extraction. We conducted experiments using five renowned pretrained networks as 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 pretrained networks was evaluated in the Results and Performance section. While VGG-19 and AlexNet exhibited acceptable performance, their large sizes led to their exclusion from consideration. SqueezeNet, being lightweight with fewer learnable parameters, resulted in decreased overall network accuracy. In comparison, GoogleNet outperformed SqueezeNet and ResNet50. The final classification accuracy for VGG-19, AlexNet, and GoogleNet was nearly identical. Therefore, considering all factors, GoogleNet was selected for feature vector extraction in Algorithm 1.

### 3) FRONT-DOOR SECURITY (FDS) ALGORITHM

Algorithm 2, implementing the FDS, utilizes the trained BiLSTM network for classifying the four actions outlined in table 1. This algorithm operates by continuously reading the real-time video stream. Upon activation of the camera, it processes the frames. When a substantial variance is detected between two consecutive frames, the algorithm engages the BiLSTM network, initiating frame transmission. The BiLSTM network then provides a prediction label (p) for the action depicted in the video, along with a confidence score (s). Utilizing the predicted label and confidence score, the algorithm employs the security application API to notify the user accordingly. Conversely, if no significant difference is detected between successive frames, the algorithm remains inactive.

## IV. RESULTS AND PERFORMANCE EVALUATION

The effectiveness of the FDS algorithm relies heavily on the accuracy of the proposed GoogleNet-BiLSTM hybrid network. Given its Deep Learning (DL) nature, we employed state-of-the-art machine learning performance evaluation metrics [48] to initially assess the performance of the proposed network. Subsequently, we conducted a comparative analysis of performance across various models, illustrating the superiority of our proposed methodology. Following this comparison, we conducted an additional experiment utilizing diverse datasets to evaluate the robustness of the system. Although the proposed model was trained using the HMDB51 dataset, we utilized videos from all datasets mentioned in Table 2 to investigate the performance of the proposed hybrid network.

### 1) OVERALL PERFORMANCE

This section presents the performance evaluation of the GoogleNet-BiLSTM hybrid network. Table 4 lists the performance evaluation metrics utilized, along with their mathematical expressions and roles in the evaluation process. These metrics, which include accuracy, sensitivity, false positive rate (FPR), and false negative rate (FNR), require the values of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). To obtain these values, we generated a confusion matrix, as illustrated in Figure 10.

Based on the values obtained from the confusion matrix and using the mathematical expressions for the evaluation metrics, the performance of the proposed methodology is detailed in Table 5. This table provides accuracy, sensitivity, FPR, and FNR values for each class individually, with average values of 73.18%, 68.25%, 69.78%, and 71.10%, respectively. Additionally, precision, recall, and F1-score for each of the four different classes were directly calculated from the confusion matrix and are listed in Table 6.

The average precision, calculated as 0.7149, attests to the quality of positive predictions made by our network. The ratio of correctly identified positive class instances to the total number of positive samples, denoted as recall, is 0.6825. Furthermore, the harmonic mean, or F1-score, is calculated as 0.6447. Statistical analysis of these numerical values confirms the acceptability of the performance exhibited by the proposed hybrid network. This assertion is further supported by a comparison study presented in the subsequent section.

### 2) PERFORMANCE COMPARISON AMONG DIFFERENT MODELS

Our experimentation involved comparing the performance of Multi-layer Perceptron (MLP), Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), and Bidirectional Long Short-Term Memory (BiLSTM) models. For each class, we randomly selected 25 video clips, which were then categorized into 30-second and 60-second durations. To ensure fairness, all models processed video sequences of equal duration.

Leveraging GoogleNet for video feature extraction contributed to a 2.80% improvement in performance. This underscores the effectiveness of our methodology in enhancing human action recognition accuracy from CCTV video frame streams. It is evident that our proposed model, a fusion of CNN and BiLSTM, outperforms other models.

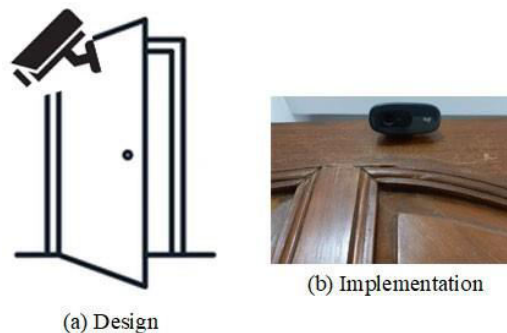
Notably, better performance was observed on shorter videos compared to longer ones, yet our proposed model consistently demonstrated superior performance in both scenarios.

### 3) PERFORMANCE COMPARISON AMONG DIFFERENT DATASETS

Before devising our methodology, we analyzed seven distinct datasets. While our proposed network was trained, tested, and evaluated using the HMDB51 dataset, we sought to further assess its performance by conducting experiments with the other seven datasets outlined in Table 2. In this experiment, we specifically selected similar classes across datasets. Subsequently, we randomly selected video clips from each class and trimmed them into 30-second and 60-second segments. For consistency, 20 video clips were utilized from each class.

## V. REAL-WORLD IMPLEMENTATION

While the performance of our proposed front door security system in laboratory experiments is deemed satisfactory, disparities between laboratory and real-world scenarios are inevitable. To bridge this gap, we deployed the security system at the front door of an apartment, as depicted in Figure 12. Utilizing a Logitech C270 HD webcam, we mounted it at the top of the door at a 45-degree viewing angle. The camera was configured to capture video at 30 frames per second (FPS). To minimize potential delays, we opted for a wired connection via USB cable instead of wireless communication. The camera is linked to a laptop computer, which hosts the installed proposed FDS algorithm.



### 1) OBSERVATION & ANALYSIS

During a 600-second period, we observed the performance as an actor randomly executed actions such as kicking, hitting, punching, and drawing a gun at the door. Human observation confirms the FDS algorithm's effective identification of these target incidents. Additionally, we conducted a numerical analysis of the proposed system's real-world performance. In this process, we captured video footage of each incident lasting 150 seconds. Subsequently, we assessed the duration for which the FDS system correctly and incorrectly identified the incidents. From the ratio of correct to incorrect identification durations, we derived a performance score.

## VI. LIMITATIONS AND FUTURE SCOPE

Every computing system has its set of limitations, and the proposed FDS is no exception. Despite demonstrating effectiveness in addressing real-world challenges, it possesses several limitations that warrant consideration.

### 1) SUBJECT-CAMERA ANGLE SENSITIVITY

One limitation of the proposed methodology pertains to subject-camera angle sensitivity. It has been noted that when the subject is positioned too close to the camera, at an angle of around 70 degrees, the accuracy of classes such as kicking and punching increases while accuracy in other classes decreases. Optimal performance is achieved when the angle ranges between 40 to 60 degrees. However, as the angle exceeds 60 degrees, performance begins to degrade. The experimental results outlined in this paper were obtained within the 40 to 60-degree range. This limitation presents an opportunity for further research to enhance the robustness of the FDS.

### 2) NIGHT-VISION EXPERIMENT

Security systems are essential for both daytime and nighttime surveillance. While our proposed methodology has been tested in daylight and under floodlights at night, conducting experiments in night-vision mode remains an unexplored aspect in this paper. However, ongoing research endeavors aim to investigate the system's performance under night-vision conditions.

This paper addresses four specific security threats at the front door, acknowledging the existence of numerous other potential risks that were not included due to data availability constraints. The creation of a custom dataset tailored for the FDS algorithm would enable the detection of all common security threats at the front door, facilitating further research avenues. This includes the development and processing of new datasets, identification of limitations, and enhancements for additional classes, as well as addressing computational complexities.

Rather than viewing the limitations of the proposed FDS algorithm as constraints, the researchers behind this project perceive them as opportunities for advancing the system. These limitations serve as catalysts for ongoing research efforts aimed at integrating innovative features and refining the algorithm.

### 3) LIMITATIONS IN ACTIVITY COVERAGE

While this paper addresses four specific security threats, it acknowledges the existence of numerous other potential risks at the front door that were not included due to data availability constraints. The creation of a custom dataset tailored for the FDS algorithm would enable the detection of all common security threats at the front door, opening avenues for further research. This includes the development and processing of new datasets, identification of limitations, and enhancements for new classes, as well as addressing computational complexities. Instead of viewing these limitations as obstacles, the researchers behind this project regard them as opportunities for advancing the system. These limitations serve as catalysts for ongoing research efforts aimed at integrating more innovative features and refining the algorithm.

## VII. CONCLUSION

Human Activity Recognition (HAR) has garnered significant attention from both the wearable sensors and computer vision research communities. In this study, we developed a hybrid network by amalgamating cutting-edge techniques prevalent in current research trends. Our innovative approach presents a promising solution for enhancing front-door security. While the advancement of HAR research is notable, its practical application in front-door security remains largely unexplored. Existing AI surveillance services, albeit effective, are often costly and require extensive infrastructure. In contrast, the FDS algorithm proposed in this paper does not necessitate additional equipment. Simply integrating CCTV camera video streams or basic webcams is sufficient to detect security threats with an accuracy of 73.1%, optimized to minimize false alarms. Real-world implementation and experimental results demonstrate the adaptability of the FDS system in fortifying front-door security. Nonetheless, acknowledging the limitations of the FDS algorithm alongside its impressive performance opens up opportunities for enhancing the system's service quality and robustness. This, in turn, aims to democratize intelligent front-door security, making it affordable and accessible to all.

## 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. Faruqi et al., "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 et al., "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 et al., "Applications of artificial intelligence in transport: An overview," *Sustainability*, vol. 11, no. 1, p. 189, 2019.
- [7] S. Laato et al., "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 et al., "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: Employing an ERNIE-BiLSTM approach for bullet screen comments," *Sensors*, vol. 22, no. 14, p. 5223, Jul. 2022.





- [12] Y. Li and L. Wang, "Human activity recognition utilizing residual network and BiLSTM," *Sensors*, vol. 22, no. 2, p. 635, Jan. 2022.
- [13] P. Chorghe, S. Pathak, and S. Joshii, "A review of intelligent surveillance systems 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, "Recognition of human activity and motion patterns within indoor environments 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, "Recognition of sports actions based on deep learning and clustering extraction algorithms," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–9, Mar. 2022.



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details