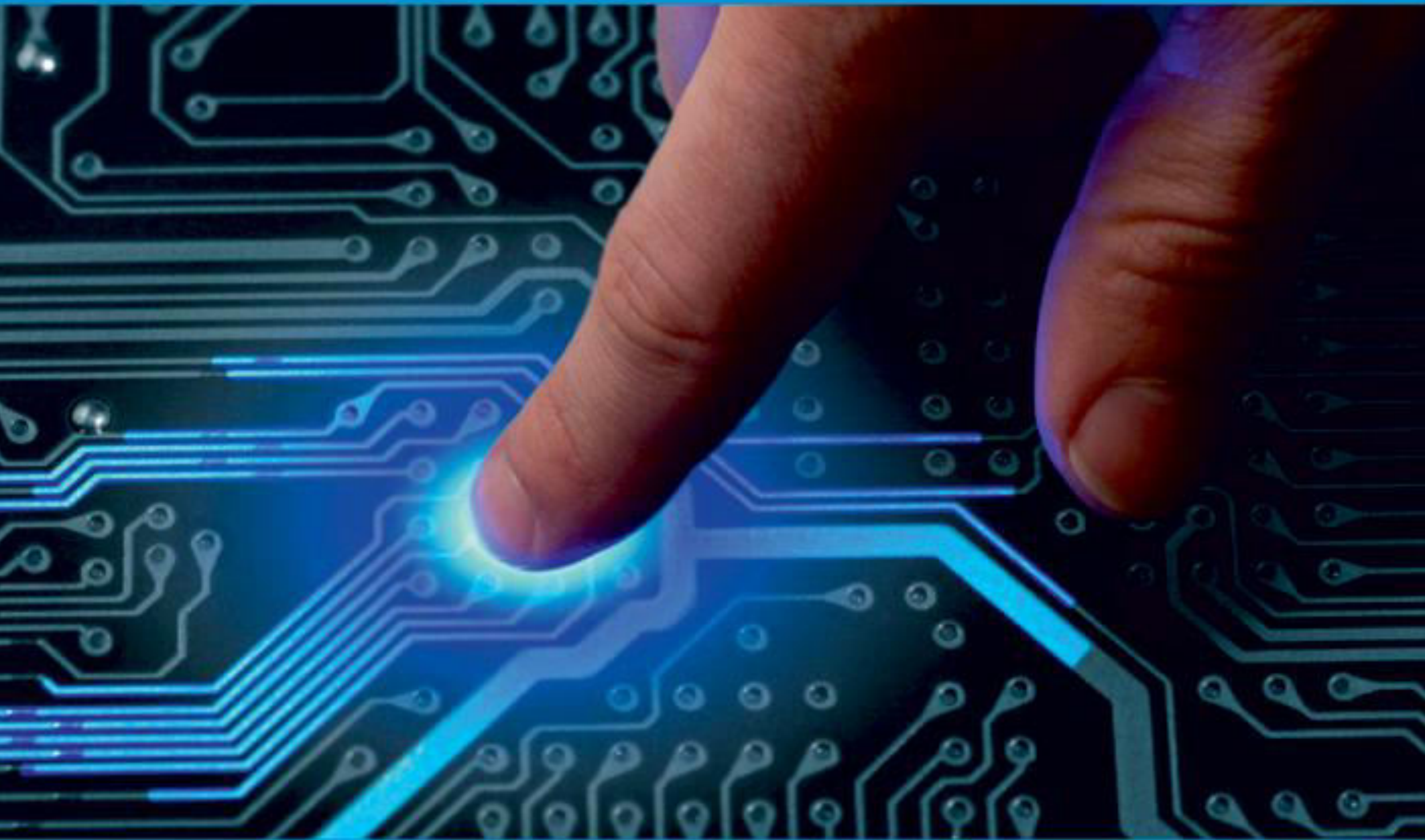
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 5, May 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105218 |

Crowd Monitoring and Misbehavioural Activity Detection using Deep Learning

Rashmi Ramteke, Chaya Jadhav

PG Student, Department of Computer Engineering, Dr. D. Y. Patil Institute Technology, Pimpri, Pune, India

Associate Professor, Department of Computer Engineering, Dr. D. Y. Patil Institute Technology, Pimpri, Pune, India

ABSTRACT—For both military and civilian real-time applications, visual surveillance is a highly sought-after solution. Crowd surveillance becomes a tiresome undertaking for large events like rallies and stadiums. Additionally, assessment of the crowd density is crucial for enhancing the effectiveness of crowd surveillance systems. For crowd counting, crowd density estimates, and crowd behavior analysis, many methodologies have been given. Noise, opacity, and a congested environment in crowd situations make it difficult to understand how people behave in them. We provide a combined approach to perform crowd density estimate and crowd behavior analysis in order to address the difficulties associated with crowd monitoring. A novel Fully Convolutional Network (FCN)-based architecture that takes into account scale information and circumvents scale fluctuations is used to estimate the crowd density. We offer a scale-aware attention module, which concatenates numerous branches of the self-attention module to better realize scale variation, to address the scale variation issue. Additionally, we provide a crowd behavior analysis algorithm that identifies aberrant crowd behavior using motion maps and energy level distribution-based characteristics. The proposed method detects anomalous events with an average accuracy of 96.45 and provides results in a timeliness manner.

KEYWORDS*:* Crowd monitoring, Deep Learning, FCN, suspicious activity, performance

I. INTRODUCTION

The usage of diverse applications, which helps us in many aspects of daily life, has recently spread like wildfire. Among these applications, visual systems are among the most important. Currently, dividing suspicious human behavior into categories for normal and abnormal behavior is a common step in studies on the topic of recognizing suspicious human behavior via video surveillance. According to [1], there are increasingly more suspicious and disruptive actions taking place in a variety of public locations throughout the world, including banks, businesses, hospitals, airports, shopping centers, exam rooms, and train stations. In order to guarantee total security, CCTV cameras are installed in the majority of organizations to prevent a variety of security difficulties. We reside in a nation with a sizable population, and as a result, many people are captured by those cameras, and numerous recordings are produced and kept for a certain period of time in various locations. After a complaint, captured recordings are examined for criminal activity. It would need a lot of workforces and human attention to continuously watch and check that the actions taking place in front of them are suspicious since continuous monitoring is a really difficult undertaking. Most activity detection systems, which employ generalized models to recognize various suspicious activities in films, suffer from a lack of performance efficiency and accuracy [2].

In order to complete this project, reliable automatic video detection must replace manual monitoring. Video surveillance is very important today. When machine learning, artificial intelligence, and deep learning are included into the systems, the technologies are created more successfully [3]. By utilizing various models and techniques of image processing, machine learning, and deep learning, a lot of work is completed, such as object detection, human behavior recognition, and human activity detection [4]. Machines can behave like humans thanks to artificial intelligence. The system will automatically detect suspicious activity and deliver an alarm when it discovers aberrant or suspicious activity in a video that may be archived or based in real-time. For the purpose of identifying and classifying these behaviors as suspicious, this study uses a deep learning model. The six acts that the proposed system would focus on include running, punching, falling, grabbing, kicking, and shooting. For the protection of various public spaces and for early crime prevention, these actions might be detected [5].

The application of multiple deep learning models and methodologies allows for the speedy detection of a variety of suspicious actions. The deep learning models are used to automatically highlight important characteristics and provide powerful representations of visual data. Convolutional Neural Network (CNN), which can learn visual patterns directly from a video frame and is best suited for that purpose, is utilized for the detection of the activities from a video using a

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105218 |

deep learning technique. CNN is employed in several types of study, including automatic threat sign identification, object detection, and human behavior detection. The CNN model may be trained to improve authenticity using a variety of architectural types, including Resent, Alex net, VGG net, and dense net [6].

In their work, [7] present a method based on digital image and video processing for identifying and tracking the movement of multiple objects in an occlusion environment, as well as for sounding an alert if an object is dropped for an extended period of time in the camera's field of view. According to this study, tasks requiring activity identification and early detection, as well as learning activity progression, benefit more from improvements in the training of deep temporal models. They discovered that while training an LSTM model for a recurrent neural network, classification error is taken into account [8]. In this study, long-term temporal convolution (LTC) neural networks are used to train video representations. They examine the impact of several low-level representations, including optical flow vector fields and randomly generated video pixel values. They present the most current findings on the difficult human activity identification datasets [9], although additional activities are still required for better performance and activity detection.

The next portions of this research report go into the literature review of current approaches, the suggested architecture for the smart irrigation research effort, the kinds of data that will be used, and the study's findings.

II. LITERATURE SURVEY

The following findings for each system were reached after the authors completed a thorough analysis of many different crowd monitoring systems that are currently in use.

Convolutional Neural Networks (CNNs) are now a recognized class of models tackling challenges with image identification, according to [10]. In this work, a new dataset of 1 million videos divided into 487 categories is used to categorize films on a large scale using CNNs. They examine several strategies for using local spatiotemporal data to strengthen a CNN's link in the time domain and suggest a multi-resolution architecture to hasten training. They investigate the best model's generalization performance and find substantial increases over the UCF-101 baseline model (63.3 percent up from 43.9 percent) after retraining the top layers using the UCF-101 Action Recognition dataset.

Recent Convolutional Neural Networks (CNN) applications for human activity recognition from movies have provided numerous algorithms that take appearance and motion into account, according to [11]. They look for numerous ways to use this spatiotemporal data to enjoy CNN towers. They discover that there is no performance loss when combining a spatial and temporal network at a convolution layer as opposed to a soft max layer. The spatial-temporal pooling of abstract convolutional features lastly improves performance. Based on this research, they suggest a unique CNN architecture for the spatiotemporal fusion of video extraction, test it under standard conditions, and show that it produces state-of-the-art outcomes.

According to [12], a recognition approach using a Convolutional Neural Network (CNN), one of the most complex algorithms composed of different layers (Convolutional layer, fully connected layer, and max pooling), is used to increase the precision of indoor human activity detection based on geographical location data. Six actions are identified by the classifier: walking, lying down, standing up, running, and leaping.

When the trial results are divided, the best identification rate of the several experimenters is 86.7 percent, proving that it is feasible.

The use of deep learning and translated learning approaches in the fall detection process using data processing from security cameras was reviewed and researched by [13]. The National Centre for Scientific Research provided an open dataset to the Laboratory of Electronics and Imaging. The CNN Alex Net architecture served as the foundation for the classifier, which was later adjusted to overcome the fall detection issue. This study evaluates automatic fall detection using data from security cameras collected under actual conditions. By adding additional algorithms that take into account the relative locations of the frames in time and our knowledge of the typical duration of fall occurrences, they hope to improve the classifier's performance.

Human action recognition has developed into a challenging video interpretation and analysis research area, according to [14]. Patient monitoring is one real-time application of human activity recognition; patients are tracked among a group of healthy individuals and afterwards identified based on their peculiar behavior. To construct a multi-class aberrant activity detection in individuals and groups, they use video sequences. The foundation of the CNN model in this study is the You Look Only Once (YOLO) network. They retrained the foundational CNN model over 32 iterations using 23,040 annotated pictures of patient behaviors. In this experiment, the action recognition accuracy was 96.8%.

The usage of an intelligent deep feature-based anomaly detection framework that may function in a surveillance network with reduced time complexity was explored by [15]. To extract data from video frames, this

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105218 |

architecture employs a Convolutional Neural Network (CNN) model that has already been trained. After that, the Bi-directional long short-term memory (BD-LSTM) model is used to analyze the frames. The UCF-Crime dataset contains records of both common and unusual events, including 13 unique oddities including fights, abuse, and accidents. The collection includes the 1900 surveillance recordings, with accuracy rates of 3.41 percent and 8.09 percent, respectively.

We identify six suspicious behaviors (Running, Punching, Falling, Snatching, Kicking, and Shooting) in the proposed work that have never been identified in a previous study. An automatic video detection system is required since it is difficult to continuously watch camera footage captured in public spaces to look for any unusual events. However, none of the intelligent surveillance systems developed by several researchers have flawless detection and accuracy rates [16].

The following shortcomings of the current crowd surveillance methods are noted from the thorough literature surveys:

1. The primary issue with the current monitoring method is people counts and estimating in packed events.
2. Recognizing and monitoring a person's movements through the crowd by using successive frames.
3. Nearly all surveillance systems struggle with accurate crowd behavior analysis.

III. OBJECTIVES

The following objectives are developed from the constraints of the current crowd monitoring technologies and the indicated research need. These are the research goals for creating a system for crowd surveillance that can reliably identify suspicious activity before it occurs and thereby shield society from unwelcome criminal activity.

1. To create a system for crowd surveillance that can identify suspicious activity.
2. To create a crowd surveillance system employing FCN and VGG16 to precisely identify and track a person's locations in congested situations.
3. To create a system that accurately analyses crowd behavior using FCN and VGG16 in order to stop shady activities in crowded regions.

The constraints in the literature study are linked to all of the aforementioned objectives.

IV. PROPOSED METHODOLOGY

Fully Convolutional network (FCN):

Without the need of additional hardware, it is demonstrated that a fully convolutional network (FCN) that has been trained end-to-end and pixel-by-pixel on semantic segmentation performs better than the state-of-the-art. This is the first research that we are aware of that fully trains FCNs for both supervised pre-training and pixel wise prediction. Existing networks' fully convolutional iterations forecast dense outputs from arbitrary-sized inputs [17]. Back propagation and dense feed forward processing are used to learn and infer one entire picture at a time. Pixel wise prediction and learning are made possible in nets with sub sampled pooling by in-network up sampling layers. This strategy is both absolutely and asymptotically effective and does not require the challenges seen in other attempts. Although patch wise training is widely employed, it performs less well than training with complete convolutions. With our approach, pre- and post-processing problems like super pixels, recommendations, and impromptu post-processing enhancement by local classifiers or random fields are avoided. By reinterpreting classification nets as completely convolutional and modifying from their learnt representations, our strategy extends prior success in classification to dense prediction.

The process is seen in fig. 1. The diagram illustrates how video frames from an incoming stream are used to create the input pictures. The VGG16 model is used to extract the necessary picture attributes from these photos in order to detect crowd behaviors and track individual people. The fully convolutional network and LSTM receives the input from the preprocessed images and analyses it for classification.

The crowd analysis with individuals in the photos, their activities based on motions, which will be analyzed by FCN along with LSTM through successive image scanning, are extracted by the classification model using FCN and LSTM, which analyses the image data. The FCN + LSTM will recognize the people, follow their movements, and record their deeds. The FCN + LSTM categorization will identify suspicious behavior when an individual or group of persons in a crowd exhibits behavior that is deemed odd and undesired. The FCN-based model evaluates the incoming video streams

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |
|| Volume 11, Issue 5, May 2023 ||
| DOI: 10.15680/IJIRCCE.2023.1105218 |

of the crowd rather than manually visualizing them, creating an intelligent surveillance system. The FCN-based model only gives an alarm if there is a likelihood of a crime or other suspicious conduct. In order to monitor the crowd for indications of criminal activity, advanced technology has replaced manual surveillance completely.

V. Data Requirements

The image data having images of events, crowds or crowd areas is the input data of the model for training and testing of the model. The dataset is available at kaggle for crowd monitoring and crowd counting activities.

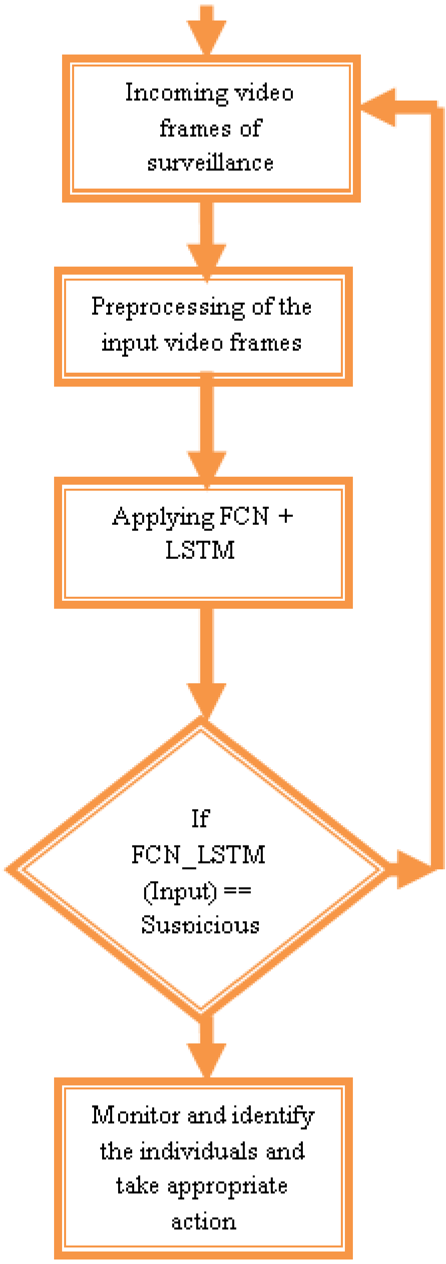


Fig. 1 Proposed Architecture

VI. Results and Discussion

Here, the FCN + LSTM model is trained and tested with the kaggle crowd monitoring and crowd counting dataset. The results are generated and accuracy of the system is tested. It has provided the accuracy of 98.85% for detecting the unwanted and suspicious activity correctly from the available video surveillance. Figure 2 shows the comparison of the model with the existing systems of crowd monitoring and crowd counting.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105218 |

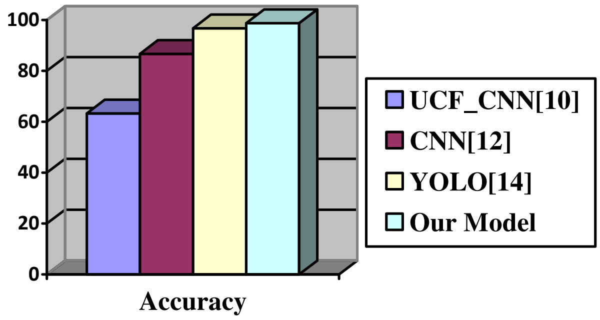


Fig. 2: Accuracy comparison

VII. CONCLUSION

This research focuses on a fully automated method for watching crowds and spotting suspicious actions. It was done to detect suspicious crowd behavior using deep learning and FCN. It describes a method that uses the VGG16 and FCN to accurately detect suspicious behavior without the need for human interaction. The government forces and security agencies' workloads are reduced by this system since they no longer have to manually analyze the video traffic originating from the system's CCTV footages in order to run it. The automation in the model will operate the alerts in case of suspicious actions in the crowd situations as per demand once it is instantiated in the real environment. Future plans for the system include the use of real-time input footages and the development of efficient deep learning techniques to improve accuracy while shortening development times.

REFERENCES

[1] Amrutha CV, Jyotsna C, Amudha J (2020) Deep learning approach for suspicious activity detection from surveillance video. In: 2020 2nd international conference on innovative mechanisms for industry applications (ICIMIA), pp 335–339. IEEE

[2] Anishchenko, L (2018) Machine learning in video surveillance for fall detection. In: 2018a ural symposium on biomedical engineering, radioelectronics and information technology (usbereit), pp 99–102. IEEE

[3] Bibi I, Akhunzada A, Malik J, Iqbal J, Musaddiq A, Kim S (2020) A dynamic DL-driven architecture to combat sophisticated android malware. IEEE Access 8:129600–129612

[4] Butt UM, Letchmunan S, Hassan FH, Zia S, Baqir A (2020) Detecting video surveillance using VGG19 convolutional neural networks. Int J Adv Comput Sci Appl 11

[5] Cheng K-W, Chen Y-T, Fang W-H (2016) An efficient subsequence search for video anomaly detection and localization. Multimed Tools Appl 75:15101–15122

[6] Desai P, Sujatha C, Chakraborty S, Ansuman S, Bhandari S, Kardiguddi S (2022). Next frame prediction using ConvLSTM. Journal of Physics: Conference Series. IOP Publishing, pp.12024

[7] Fan Y, Wen G, Li D, Qiu S, Levine MD, Xiao F (2020) Video anomaly detection and localization via gaussian mixture fully convolutional variational autoencoder. Comput Vis Image Underst 195:102920

[8] Feichtenhofer C, Pinz A, Zisserman A (2016a) Convolutional twostream network fusion for video action recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1933–1941

[9] Gul MA, Yousaf MH, Nawaz S, Ur Rehman Z, Kim H (2020) Patient monitoring by abnormal human activity recognition based on CNN architecture. Electronics 9:1993

[10] Htike KK, Khalifa OO, Ramli HAM, Abushariah MAM (2014) Human activity recognition for video surveillance using sequences of postures. In: The third international conference on e-technologies and networks for development (ICeND2014), pp.79–82. IEEE

[11] Joshi K, Tripathi V, Bose C, Bhardwaj C (2020) Robust sports image classification using InceptionV3 and neural networks. Procedia Comput Sci 167:2374–2381

[12] Karpathy A, Toderici G, Shetty S, Leung T, Sukthankar R, Fei-Fei L (2014). Large-scale video classification with convolutional neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1725–1732

[13] Li J, Wu R, Zhao J, Ma Y (2017) Convolutional neural networks (CNN) for indoor human activity recognition using Ubisense system. In: 2017 29th Chinese control and decision conference (CCDC), pp 2068–2072. IEEE

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105218 |

[14] Ma S, Sigal L, Sclaroff S (2016) Learning activity progression in lstms for activity detection and early detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1942–1950

[15] Mahdi MS, Mohammed AJ (2021) Detection of unusual activity in surveillance video scenes based on deep learning strategies. J Al-Qadisiyah Comput Sci Math 13:1

[16] Perez M, Kot AC, Rocha A (2019) Detection of real-world fights in surveillance videos. In: ICASSP 2019–2019 IEEE international conference on acoustics, speech and signal processing (ICASSP), pp 2662–2666 IEEE

[17] Sonkar, Riddhi & Rathod, Sadhana & Jadhav, Renuka & Patil, Deepali. (2020). CROWD ABNORMAL BEHAVIOUR DETECTION USING DEEP LEARNING. ITM Web of Conferences. 32. 03040. 10.1051/itmconf/20203203040.

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.379

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर
NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com