Abnormal Action Recognition from Surveillance Visual using Deep learning

[1]Meera Raghuvanshi, [2]Manpreet Kaur, [3]Jatin Rajput , [4]Harsh Tomar

[1,2,3,4] Department of Computer Science and Technology Meerut Institute of engineering and technology, Meerut

meera.raghuvanshi.cs.2019@miet.ac.in , manpreet.kaur.cs.2019@miet.ac.in ,
jatin.rajput.cs.2019@miet.ac.in , harsh.jitendra.cs.2019@miet.ac.in

ABSTRACT

In the modern-day, automatic surveillance is an active research issue. It has taken the command in many fields and many remain untouched. Our main purpose is to detect anomalies at different places using CCTV. The importance of detecting suspicious human activity via video surveillance is to stop theft cases. For detecting abnormal actions we will be using Convolutional Neural Network(CNN), Neural Learning, System Learning, Long short-term memory (LSTM) and Deep Structured Learning for detection of abnormal actions of students on campus. The use of artificial intelligence enables computers to think like people. Making predictions based on future data and learning from training data are crucial aspects of machine learning. Further, Deep learning is employed since there are now GPU (Graphics Processing Unit) processors and large datasets accessible. In previous research LSTM was used which was not so effective, therefore Long-term recurrent convolutional network (LRCN) training model will be used here to increase the accuracy of the anomaly detection. The proposed model will be able to detect human normal and suspicious activities based on the training model.

Keywords- Recognition, Neural Learning, System Learning, Deep Structured Learning, Long short-term memory, Long-term recurrent convolutional network

INTRODUCTION

Security Camera Supervision is now the most reliable measure a location can contain. It is something that we are likely to find everywhere, including hospitals, city malls, universities, etc. But picture a campus with over a hundred CCTVs spread across various buildings, hostels, classrooms, canteens, sports fields, auditoriums, etc. Today, it is difficult to manually watch every incident. Even if the incident has already happened, manually looking for it in the captured video is a time-consuming process .One of the greatest designs for tackling challenging learning tasks is deep neural networks. Automated feature extraction in Deep Learning models enables the creation of generalized, high-level representations of picture data. Convolutional neural networks (CNN) can directly learn visual patterns from picture pixels, while Long short-term memory (LSTM) models can detect long-term dependencies within video streams by retaining information. By combining these two models, a Long-term recurrent convolutional network (LRCN) will be created to monitor CCTV footage and identify any unusual activities like — running, walking & fighting. The system can be used to generate a notification that will notify the user if any uncertain activity is detected. Human Actions are under observation via different intelligent models at Railway and Bus stations, Airports, Schools, Stadiums, Offices, Hospitals, Parking Area, Malls and many more to prevent theft, attacks, Snatching, crime and other abnormal activity. The suggested method will employ CCTV camera video to monitor campus activity and gently alert users when anything unusual happens. Intelligent video surveillance relies on identifying events and recognizing human behavior, as these are the essential components of the system.[1][2][3].

Three phases can be used to denote the whole surveillance system training process: Preparation of the data, Model training, and Inference

LITERATURE REVIEW

The associated papers provide several methods for identifying human activities in video.The works' main goal was to find any unusual or suspicious activities in a security film. The advanced motion detection (AMD) technique was used to detect an unauthorized entry into a confined space. The object was identified by initially applying background subtraction, and then further analyzed using structure progression.The arrival of questionable activities was the second stage. The algorithm used by the system has the advantage of real-time video processing and minimal computing complexity. However, the system had storage service limitations and may also be used to construct a high-tech method of video collecting in monitoring zones [4] [5][6].

In order to prevent violent behavior in crowds, a real-time violence detection system was developed using deep learning technology. The system captures frames from real-time videos in a Spark environment and immediately alerts security personnel if it detects any signs of violence. By continuously monitoring the video feed, the system is able to recognize potentially violent actions in real-time and provide early warning to security officers to prevent violent incidents from occurring. To identify individuals and unusual incidents in the video footage, a background subtraction technique is used to locate human figures. The features extracted using a CNN are then inputted to a DDBN (Discriminative Deep Belief Network) for analysis [2][7][8][9].

The authors demonstrate that their technique beats some of the previously reported findings by up to 9% on the opportunity dataset by using a mix of deep convolutional networks and LSTM to conduct multi-modal wearable activity identification. The authors used time-series data gathered from smartphone sensors together with neural networks to categorize activities. According to experiments, adding more convolutional layers improves performance, but each new layer reduces the complexity of the produced features [3][10][11].

Before being analyzed by the DDBN system, videos depicting questionable incidents are labeled and relevant characteristics are extracted from them. The DDBN then compares the features obtained from labeled sample videos of categorized suspicious behaviors to features obtained using a CNN, resulting in the identification of several suspicious activities within the provided video footage. To prevent violence by the audience or players during sporting events, a real-time violence detection system that uses deep learning technology was developed. Frames are extracted from real-time videos in a Spark environment, and alerts are sent to authorities if the system detects any instances of football-related violence [12][13].

PROPOSED METHODOLOGY

In our suggested system, the LRCN (Long-term Recurrent Convolutional Network) has been employed to identify unexpected actions. Recognizing the temporal data in the video is vital for accurate categorization of odd behaviors. CNN has recently been utilized mostly for extracting important characteristics from each video frame. The features must be taken from CNN in order to effectively categorize the input, hence CNN must be able to recognise and extract the required characteristics from the video frame. The videos were captured to prepare a dataset. Each video is preprocessed in 3 forms: video clipping, frame extraction from the videos and dataset cleaning. Once the cleaning is done then processed data is passed further to extract keyframes from the data. The video's multiple frame sequence is taken out and passed through the CNN algorithm and sent to the LRCN Model. Based on which further predictions were made. The suggested method will employ CCTV camera video to record student activity on a campus and alert the appropriate authority when something suspect happens. The design is divided into several stages, including video capturing, video pre-processing, key-frame extraction, Algorithm and class prediction. The method divides the videos into three groups, as follows [14][15][16].

- On-campus fighting - Suspicious class
- Strolling and running - Regular class

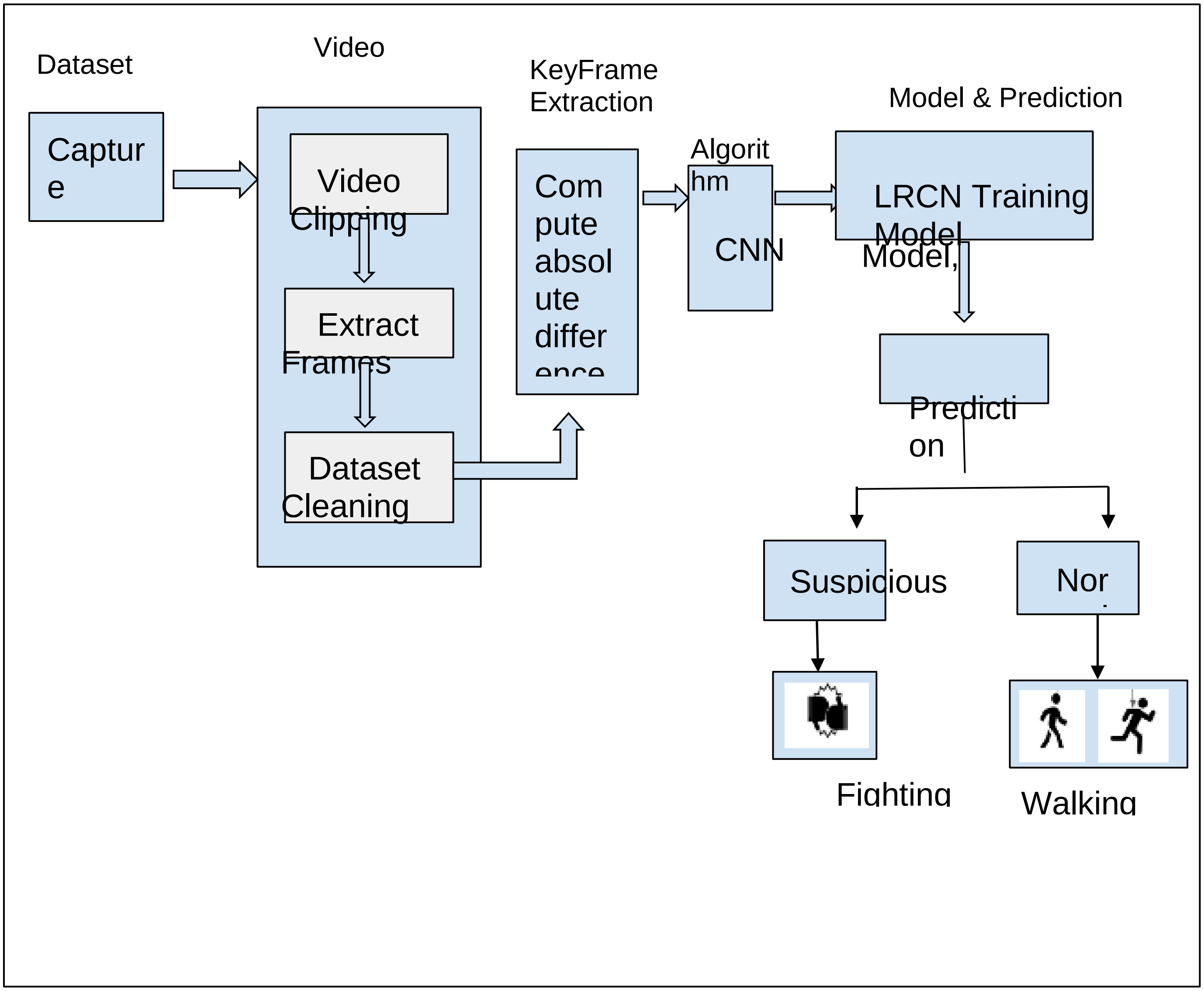


Figure 1: The proposed model architecture of LRCN

Datasets

The KTH dataset is utilized for recognizing walking and running activities, while the Kaggle dataset is specifically designed for detecting fighting. The KTH dataset contains a set of 6 tasks with 100 series for each action type, and each segment includes more than 600 frames captured at a rate of 25 frames per second. On the other hand, the Kaggle dataset contains over 100 footage from movies and YouTube videos, which can be used to train a model to detect suspicious behavior, specifically fighting [17][18].

Data Pre-processing

The videos are read from their respective Class folders using the OpenCV Library, and the Class label is kept within a numpy array.Dividing a video into frames to create a single sequence: Each

video is read using the OpenCV library, and only 30 frames are read at uniform intervals to create a 30-frame sequence.Resizing: When the total number of pixels has to be increased or decreased, image resizing is required. In order to retain the consistency of the input photos with the architecture, we scaled all of the frames to 64 pixels in width and height [19][20][21].
Normalization: Normalization will enable the learning algorithm to pick up crucial features from the photos more quickly. Therefore, we divided the enlarged frame to normalize it.Store in Numpy Arrays: To provide the model with input, the series of 30 scaled and normalized frames is saved in a numpy array.

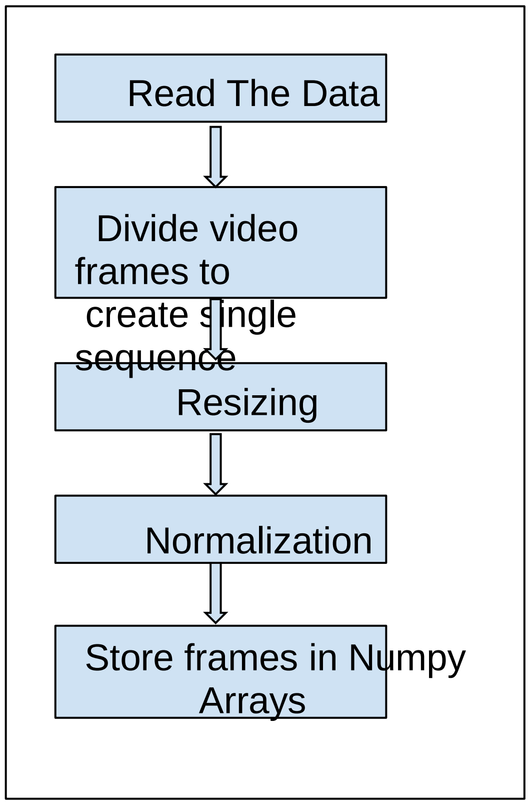


Figure 2 : The flowchart for the video preprocessing

Train Test Split Data

A fraction of 5% of the data is allocated for testing purposes, and the remaining majority of 75% is employed for training.The programme has been taught to forecast events in three classes: walking, running, and fighting. The model is trained using the training set and the following hyperparameters: Epochs equals 70, Input batch size: 4, Validation split is set at 0.25 [22][23][24].

RESULT

On our generated data set, the system achieves an accuracy of 82% in detecting the aberrant behavior occurring in the movie according to our provided model. We used the VGG-16 model, which has 16 layers, in our prior model. Consequently, it took a lot of time and could not be utilized for REAL-TIME detection. However, the LRCN model reduced the number of layers to 11 and made it faster and capable of REAL-TIME detection. To conserve memory, we downsized our frames from 224 pixels to 64 pixels, and we increased the number of films in our dataset to improve accuracy. Videos depicting unusual behavior, including fighting, are included in the dataset for the suggested model [25][26][27].

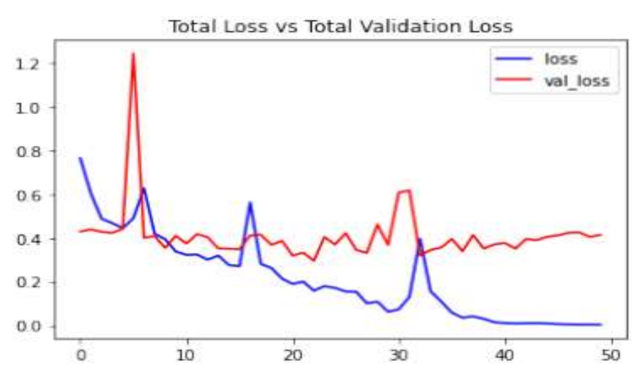
<u>Model Training</u>

The dataset is used to train the LRCN model, and a portion of 25% of the data is reserved for testing, while the majority of 75% is employed for training the model [28][29][30].

```
In [17]: # Create an Instance of Early Stopping Callback.
         early_stopping_callback = EarlyStopping(monitor = 'accuracy', patience = 10, mode = 'max', restore_best_weights = Tru

         # Compile the model and specify loss function, optimizer and metrics to the model.
         model.compile(loss = 'categorical_crossentropy', optimizer = 'Adam', metrics = ["accuracy"])

         # Start training the model.
         model_training_history = model.fit(x = features_train, y = labels_train, epochs = 70, batch_size = 4 , shuffle = True

curacy: 0.8772
Epoch 45/70
42/42 [==============================] - 1s 31ms/step - loss: 0.0085 - accuracy: 1.0000 - val_loss: 0.4053 - val_ac
curacy: 0.8947
Epoch 46/70
42/42 [==============================] - 1s 32ms/step - loss: 0.0061 - accuracy: 1.0000 - val_loss: 0.4113 - val_ac
curacy: 0.8772
Epoch 47/70
42/42 [==============================] - 1s 32ms/step - loss: 0.0050 - accuracy: 1.0000 - val_loss: 0.4235 - val_ac
curacy: 0.8772
Epoch 48/70
42/42 [==============================] - 1s 32ms/step - loss: 0.0043 - accuracy: 1.0000 - val_loss: 0.4252 - val_ac
curacy: 0.8772
Epoch 49/70
42/42 [==============================] - 1s 31ms/step - loss: 0.0040 - accuracy: 1.0000 - val_loss: 0.4044 - val_ac
curacy: 0.8947
Epoch 50/70
42/42 [==============================] - 1s 32ms/step - loss: 0.0037 - accuracy: 1.0000 - val_loss: 0.4138 - val_ac
curacy: 0.8947
```

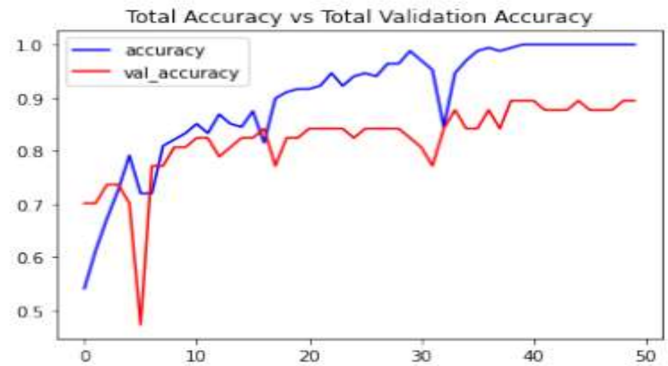
<u>Model Training graphs :</u>

(i) <u>*Loss vs Validation Loss*</u>



The graph (i) shows the total loss vs total validation loss. Under which validation loss is higher than the loss [31][32][33].

(ii) *Accuracy vs Validation Accuracy :*



The graph 2.2 shows that the model achieved an accuracy of about 82%, which is not bad for 300 records of data [34][35] [36][37].

Output Window Screen

CONCLUSION

We developed an LRCN model for the detection purpose of various activities like fighting, walking & running from Security Camera Recordings, the model was trained on 300 videos and achieved an accuracy of about 82.66%.

The paper proposes a computerized method for detecting anomalies through robust computer vision techniques using a visual attention zone to accurately locate the Region of Interest (ROI) with the aid of strong BG subtraction. The proposed model achieves 82.66% accuracy in detecting different kinds of events, such as road accidents, robbery, fighting, and running. The training network is weakly-supervised, and the method of extracting the visual attention zone is expected to be useful for various online activities, including video object localization and classification. It is worth noting that the proposed method is limited to detecting anomalous events that involve moving objects

REFERENCES

[1] Musale, Jitendra, Akshata Gavhane, Liyakat Shaikh, Pournima Hagwane, and Snehalata Tadge. "Suspicious Movement Detection and Tracking of Human Behavior and Objects with Fire Detection using Closed Circuit TV (CCTV) cameras." International Journal for Research in Applied Science & Engineering Technology (IJRASET) Volume 5 (2017).

[2] Divya, P. Bhagya, S. Shalini, R. Deepa, and Baddeli Sravya Reddy. "Inspection of suspicious human activity in the crowdsourced areas captured in surveillance cameras." International Research Journal of Engineering and Technology (IRJET) (2017).

[3] Arifoglu, Damla, and Abdelhamid Bouchachia. "Activity recognition and abnormal behavior detection with recurrent neural networks." Procedia Computer Science 110 (2017): 86-93.

[4].K.Kumar., " Comprehensive Method of Botnet Detection Using Machine Learning", International Journal of Open Source Software and Processes (IJOSSP), 12(4), 1-25,2021.

[5].K,Kumar., &V, Bhatnagar., "Machine Learning Algorithms Performance Evaluation for Intrusion Detection", Journal of Information Technology Management, 13(1), 42-61,2020.

[6].K,Kumar., &V, Bhatnagar, "Credit Card Fraud Transaction Detection System Using Neural Network-Based Sequence Classification Technique", International Journal of Open Source Software and Processes (IJOSSP), 12(1), 21-40,2021.

[7],K, Kumar., A,Kumar., V, Kumar., & S, Kumar., "A Hybrid Classification Technique for Enhancing the Effectiveness of Intrusion Detection Systems Using Machine Learning", International Journal of Organizational and Collective Intelligence (IJOCI), 12(1), 1-18,2022.

[8].K,Kumar.,"Comprehensive Composition to Spot Intrusions by Optimized Gaussian Kernel SVM", International Journal of Knowledge-Based Organizations (IJKBO), 12(1), 1-27,2022.

[9].K,Kumar., & K, S, Mishra., " Prognostication of Sales by Auto Encoder and Long-Term Short Memory", International Journal of Knowledge-Based Organizations (IJKBO), 12(1), 1-15, 2022.

[10].S,Sachdeva., T.P, Tyagi., G,Yadav., V, Kumar., & K, Kumar.,"Travel Log Analysis Tracking System Based On Salesforce For Covid-19", In Fifth International Conference on Computational Intelligence and Communication Technologies (CCICT) (pp. 358-364). IEEE,2022.

[11].K, Kumar., M, Khari., & A,Kumar.,"Comprehensive Method of Distributed Denial of Service Detection by Artificial Neural Network". In Wireless Communication with Artificial Intelligence (pp. 27-43). CRC Press,2022

[12].K.Kumar., &V. Bhatnagar, "Botnet Forensic Analytics for Investigation of Disruptive Botherders", In Disruptive Technologies for Society 5.0 (pp. 265-276). CRC Press,2021.

[13].K.Kumar., &V. Bhatnagar.,"Computing Technologies for Prognosticating the Emanation of Carbon Using an ARIMA Model", In Cyber-Physical, IoT, and Autonomous Systems in Industry 4.0 (pp. 283-295). CRC Press,2021.

[14]K. Kumar & M. Khari., "Architecture of Digital Twin for Network Forensic Analysis Using Nmap and Wireshark", In Digital Twin Technology (pp. 83-104). CRC Press,2021.

[15].K.Kumar., M.Kumar., & H.Dueby., "Blockchain in Administration: Advances and Implications of Decentralized Applications for Sharing Information", In Handbook of Green Computing and Blockchain Technologies (pp. 45-53). CRC Press,2022.

[16].S, Jain., S,Verma., H,Gupta., H., S,Jindal., S., M,Rawat.,& K,Kumar.,"Machine Learning Algorithm Based Emotion Detection System. In 2022 Fifth International Conference on Computational Intelligence and Communication Technologies (CCICT) (pp. 270-274). IEEE.

[17] Vipul, Narayan, and A. K. Daniel. "A novel protocol for detection and optimization of overlapping coverage in wireless sensor network." International Journal of Engineering and Advanced Technology 8.6 (2019): 422-462.

[18] Narayan, Vipul, A. K. Daniel, and Ashok Kumar Rai. "Energy efficient two tier cluster based protocol for wireless sensor network." 2020 international conference on electrical and electronics engineering (ICE3). IEEE, 2020.

[19] Narayan, Vipul, and A. K. Daniel. "Multi-tier cluster based smart farming using wireless sensor network." 2020 5th international conference on computing, communication and security (ICCCS). IEEE, 2020.

[20] Narayan, Vipul, and A. K. Daniel. "A novel approach for cluster head selection using trust function in WSN." Scalable Computing: Practice and Experience 22.1 (2021): 1-13.

[21] Narayan, Vipul, and A. K. Daniel. "RBCHS: Region-based cluster head selection protocol in wireless sensor network." Proceedings of Integrated Intelligence Enable Networks and Computing: IIENC 2020. Springer Singapore, 2021.

[22] Narayan, Vipul, and A. K. Daniel. "CHOP: Maximum coverage optimization and resolve hole healing problem using sleep and wake-up technique for WSN." ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal 11.2 (2022): 159-178.

[23] Narayan, Vipul, and A. K. Daniel. "CHHP: coverage optimization and hole healing protocol using sleep and wake-up concept for wireless sensor network." International Journal of System Assurance Engineering and Management 13.Suppl 1 (2022): 546-556.

[24] Narayan, Vipul, and A. K. Daniel. "IOT based sensor monitoring system for smart complex and shopping malls." Mobile Networks and Management: 11th EAI International Conference, MONAMI 2021, Virtual Event, October 27-29, 2021, Proceedings. Cham: Springer International Publishing, 2022.

[25] Narayan, Vipul, and A. K. Daniel. "FBCHS: Fuzzy Based Cluster Head Selection Protocol to Enhance Network Lifetime of WSN." ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal 11.3 (2022): 285-307.

[26] Narayan, Vipul, and A. K. Daniel. "Energy Efficient Protocol for Lifetime Prediction of Wireless Sensor Network using Multivariate Polynomial Regression Model." Journal of Scientific & Industrial Research 81.12 (2022): 1297-1309.

[27] Tyagi, Lalit Kumar, et al. "Energy Efficient Routing Protocol Using Next Cluster Head Selection Process In Two-Level Hierarchy For Wireless Sensor Network." Journal of Pharmaceutical Negative Results (2023): 665-676.

[28] Narayan, Vipul, et al. "E-Commerce recommendation method based on collaborative filtering technology." International Journal of Current Engineering and Technology 7.3 (2017): 974-982.

[29] Vipul, Narayan, and A. K. Daniel. "A novel protocol for detection and optimization of overlapping coverage in wireless sensor network." International Journal of Engineering and Advanced Technology 8.6 (2019): 422-462.

[30] Narayan, Vipul, et al. "To Implement a Web Page using Thread in Java." (2017).

[31] Srivastava, Swapnita, and P. K. Singh. "HCIP: Hybrid Short Long History Table-based Cache Instruction Prefetcher." International Journal of Next-Generation Computing 13.3 (2022).

[32] Srivastava, Swapnita, and P. K. Singh. "Proof of Optimality based on Greedy Algorithm for Offline Cache Replacement Algorithm." International Journal of Next-Generation Computing 13.3 (2022).

[33] Srivastava, Swapnita, and P. K. Singh. "Proof of Optimality based on Greedy Algorithm for Offline Cache Replacement Algorithm." International Journal of Next-Generation Computing 13.3 (2022).

[34] Pentyala, S., Liu, M., & Dreyer, M. (2019). Multi-task networks with universe, group, and task feature learning. arXiv preprint arXiv:1907.01791.

[35] Narayan, Vipul, et al. "Deep Learning Approaches for Human Gait Recognition: A Review." 2023 International Conference on Artificial Intelligence and Smart Communication (AISC). IEEE, 2023.

[36] Narayan, Vipul, et al. "FuzzyNet: Medical Image Classification based on GLCM Texture Feature." 2023 International Conference on Artificial Intelligence and Smart Communication (AISC). IEEE, 2023.

[37] Mall, Pawan Kumar, et al. "FuzzyNet-Based Modelling Smart Traffic System in Smart Cities Using Deep Learning Models." Handbook of Research on Data-Driven Mathematical Modeling in Smart Cities. IGI Global, 2023. 76-95.