Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

A Novel Convolutional Cross-Contextual Causal Dilated Neural Network with Boosted Sooty Tern Optimization based Human Activity Recognition for Suspicious Actions Detection in Surveillance Footage

Aswathy M.P[1], Dr. K.Vimalanathan[2], Dr. G. Uma Gowri[3], Vidhyashree B[4], Anitha Jaganathan[5]

[1]Assistant professor, Department of Computer science and engineering, Meenakshi college of Engineering, Chennai (0009-0001-2931-2601).
[2]Assistant Professor, Department of Industrial Engineering, College of Engineering Guindy, Anna University, Chennai 600 025 (0000-0002-9224-2044).
[3]Professor and Principal, T.S. Srinivasan polytechnic College and Advanced Training, Vanagaram, Chennai-600 095 (0000-0001-6003-4411).
[4]Assistant Professor, Computer science and Engineering (Data Science) Madanapalle institute of technology & science, Kadiri Road Angallu, Village, Madanapalle, Andhra Pradesh 517325 (0000-0003-0679-0966).
[5]Assistant Professor, Department of Artificial Intelligence and Data Science, Panimalar Engineering College, Chennai (0009-0002-7773-0469).
Email: achu.aneeee@gmail.com, anitha@panimalar.ac.in, saivimal@annauniv.edu, umagowri.g@gmail.com, vidhyashreeb@mits.ac.in

In order to detect suspicious activity in surveillance footage, Human Activity Recognition (HAR) identifies and interprets human behaviors that can point to possible threats or unusual activity. The monitoring of human activity in public and sensitive areas, like bus stops, train stations, airports, banks, shopping centers, schools, colleges, parking lots, and roadways, is made possible by visual surveillance. The purpose of this monitoring is to stop crimes and suspicious behaviors such as fighting, terrorist activity, stealing, accidents, illegal parking, criminal activity, and chain snatching. Continuous monitoring of public spaces is challenging, making intelligent video surveillance systems essential. These systems can monitor human activities in real-time, classify them as usual or unusual, and generate alerts for potential threats. The lot of techniques is implemented to recognize human activities. But the existing methods have lot of disadvantages such as high error rate and low accuracy. To

overcome the before mentioned problem, Convolutional Cross-Contextual Causal Dilated Neural Network Using Boosted Sooty Tern Optimization(4CDNN-BSTO) is proposed for recognizing human activity with high accuracy. In this input data is taken from KTH dataset. . To reduce noise in the input data, Color Wiener Filtering (CWF) is proposed. Following that, the pre-processed images undergo segmentation using Cell Attention Networks with U-Net (CAN-U-net).After that segmentation, the classification is done using Convolutional Cross-Contextual Causal Dilated Neural Network (4CDNN) and optimized using Boosted Sooty Tern Optimization (BSTO) for human activity recognition demonstrating superior efficiency and accuracy. The efficiency of the proposed 4CDNN-BSTO is analyzed using KTH dataset and attains 99.8% accuracy, 99.7 % recall and attains better results in comparison with the existing techniques. The outcomes of the proposed technique showed that it could improve the evaluations ability of the computerized human activity recognition method.

Keywords: Human Activity Recognition (HAR), Color Wiener Filtering (CWF), Convolutional Cross-Contextual Causal Dilated Neural Network (4CDNN), Boosted Sooty Tern Optimization (BSTO).

1. Introduction

Surveillance is the act of keeping an eye on behavior, activities, or other changing data, usually pertaining to people or locations in order to supervise, coordinate, or guarantee them[1]. Human Involvement Numerous real-world uses of recognition exist, such as intelligent video monitoring and analysis of consumer purchasing patterns. Video surveillance has a wide range of applications, both indoors and outside. Keeping an eye out is crucial to security maintenance. Security and safety concerns have made security cameras an indispensable part of contemporary living. E-surveillance is one of the main objectives of the Indian the government's development initiative, Digital India. It still includes video surveillance [2-4]. An automated method of intelligently identifying any suspicious action in a video monitoring system is human behavior detection. Many effective algorithms are available for automatically identifying human beings in public spaces such as banks, workplaces, exam rooms, airports, and train stations. The use of deep learning, machine learning, and artificial intelligence in the field of video surveillance is new. Artificial intelligence facilitates human-like thought processes in computers. Two essential components of machine learning are forecasting future data and learning from training data. Large data sets and GPU computers are widely available these days, which is why deep learning is used [5-7].In the current era of rising crime rates, visual monitoring is becoming increasingly important for safeguarding individuals and their belongings. Many surveillance systems are already deployed around us since excellent cameras are now widely available, but there is not enough staff to monitor the constant activity that occurs seven days a week. Additionally, a lot of footage is produced by such monitoring systems, increasing the need for storage [8-10]. This need for storage may result in higher costs. Surveillance cameras can be more valuable than passive footage recording in identifying incidents and taking immediate action. Closed-circuit television systems use thermal cameras for night-time surveillance [11-13]. Live data streaming is necessary for real-time surveillance, influenced

by factors like scene, motion, video sensor characteristics, and light.

Novelty and Contribution

- The CWF goal is to reduce noise and enhance the standard of the input data. Color Wiener Filtering (CWF) on pre-processed helps to capture detailed features better and makes the analysis of data more reliable against noise and distortions.
- The objective is to segment the pre-processed images using Cell Attention Networks with U-Net (CAN-U-net) to accurately identify and separate different cellular structures within the images.
- The objective is to classify the segmented images using a Convolutional Cross-Contextual Causal Dilated Neural Network (4CDNN) to accurately identify and categorize different features or patterns within the images. To optimize the classification process using Boosted Sooty Tern Optimization (BSTO) for human activity recognition, achieving superior efficiency and accuracy in identifying and categorizing activities, this refines the model parameters and minimizes errors, leading to superior classification results.

2. Literature Survey

In 2024, Nagalakshmi, et al [14] has introduced a Discriminative Deep Belief Network (DDBN) approach focuses on identifying suspicious human actions. It uses a convolutional neural network to extract features, converts films into frames, uses background subtraction to recognize humans, and compares the extracted features with labeled videos. When the final dataset is compared to labeled samples, the suggested classification framework performs 90% more accurately.

In 2024, Wani, et al [15] has developed an Efficient and Accurate Suspicious Activity Detection (EASAD) on Internet of Things devices with limited resources. The model achieves optimum resource efficiency without sacrificing accuracy by integrating an upgraded U-Net segmentation procedure with an updated Net architecture. To reduce computational load, the model makes use of selected feature extraction methods as SLBT, BoVW, and MoBSIFT. With 95% accuracy, the model diagnoses suspicious actions, underscoring the requirement for solutions that strike a balance between accuracy and efficiency.

In 2024, Kajendran, et al [16] has developed a Dual-Channel Capsule Generative Adversarial Network (DCCGAN) for real-time detection of human activity in public places such as bank ATMs, a supervised deep learning technique utilizing an RGB + D sensor. The Deep Convolutional Spiking Neural Network is utilized in the procedure to extract features after super pixel motion detection has identified the region of interest. Motion patterns trained on DCCGANs are created using the RGB + D database in order to detect suspicious events.

In 2024, Kersten, et al [17] has introduced a Deep learning (DL) based method. The Smart Cities development is the important component of detection for throwing activities in surveillance videos. A public dataset comprising 271 movies of thrown actions and 130 regular videos devoid of throwing movements is used in the solution. By using the optimizer

developed by Adam and putting forth a mean standard loss function that accounts for a range of traffic scenarios and reduces false alarm rates, the algorithm performs better. The ROC curve with a value of 86.10 for the Throwing-Action dataset & 80.13 for the combined dataset is displayed in the experimental results.

Problem statement:

The existing Human Activity Recognition for Suspicious Actions Detection in Surveillance Footage methods suffer from high error rate and low accuracy. To solve these issues Convolutional Cross-Contextual Causal Dilated Neural Network (4CDNN) is proposed. By employing Human Activity Recognition from KTH dataset, it enhances detection through Color Wiener Filtering (CWF) for noise reduction, and Cell attention networks with U net (CAN-U-net) for Segmentation. Convolutional Cross-Contextual Causal Dilated Neural Network (4CDNN) for classification and Optimized with BSTO, the method achieves 99.8% accuracy and 99.7% recall, outperforming current practices and maybe improving computerized Human Activity Recognition for Suspicious Actions Detection in Surveillance Footage.

3. Proposed Methodology

The working principle of 4CDNN is illustrated in Fig 1. The proposed method consists of four process (1) data acquisition, (2) pre-processing and (3) segmentation and (4) classification.

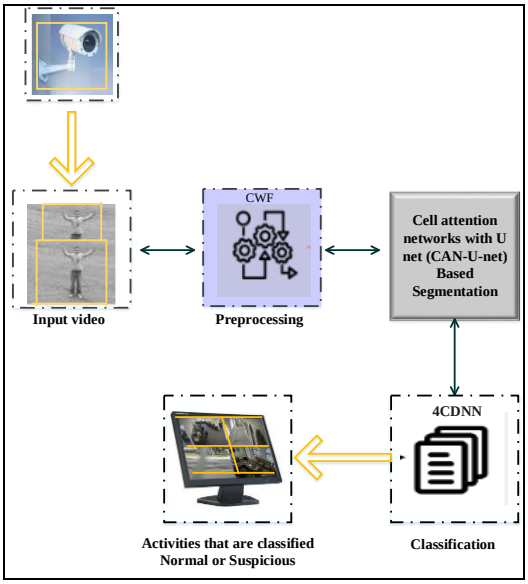

Figure 1: Workflow diagram of proposed 4CDNN method

3.1 Data Acquisition

In this input surveillance footages are taken from KTH dataset. These consists of unwanted noises, to remove these noises pre-processing techniques are used, they are given below:

3.2 Color Wiener Filtering (CWF) Based Pre-processing

Human Activity Recognition (HAR) for detecting suspicious actions in surveillance footage is a critical task in security and safety systems. Preprocessing is a crucial step in improving the accuracy and efficiency of HAR systems. Input surveillance footages are full of noise so Color Wiener Filtering (CWF) is used to remove the noises. This method is employed to improve the surveillance footages quality. In order to remove any unwanted noise, the input surveillance footages are first pre-processed. Color Wiener Filtering is a powerful preprocessing technique used in Human Activity Recognition (HAR) for suspicious actions detection in surveillance footage. It helps enhance the quality of the video data by reducing noise and preserving important details, it is given by equation (1):

$$b_j = S_j + \lambda_j \tag{1}$$

where, b_j shows the reconstructed or seen image that corresponds to the input image vector picture, S_J and λ_J suggests a noise additive .a shows the image pixel that can be expressed as a (x, y) where (x, y) are the pixel axis's horizontal and vertical directions. $\bar{b}$ Is the recovered images mean vector b_j . $\bar{S}$ is the repaired image's pixel vector and $\bar{\bar{S}}$ mean vector of the rectified image, these parameters are taken into account, and CWF filtering S is displayed by utilizing by using equation(2).

$$\hat{S} - \bar{\bar{S}} = (b_j - \bar{b}) \tag{2}$$

where, S is a mean square error (MSE) between the original and restored images must be computed. MSE is a mathematical derived by equation (3).

$$MSE = \min\left|\left(\hat{S} - \bar{\bar{S}} - R\left(b_j - \bar{b}\right)\right)^2\right| \tag{3}$$

If there is no association between the original image and noise, the observed image S is derived by equation (4).

$$S = D_{co}\left(D_{co} + D_{mm}\right) \tag{4}$$

where, D_{mm} is noisy and D_{co} is represented by its covariance matrices, which shows equation(5):

$$D_{co} = \left\langle \left(R_i - \bar{\bar{R}}\right)\left(R_i - \bar{\bar{R}}\right)^T \right\rangle \tag{5}$$

Finally, Wiener filtering S can be expressed in equation (6):

$$S = \left(D_{ee} - D_{mm}\right)D_{ee}^{-1} \tag{6}$$

Color Wiener Filtering is a valuable preprocessing technique in the HAR for suspicious actions detection in surveillance footage. By reducing noise and preserving important details, it enhances the overall performance of the HAR system, leading to more accurate and

reliable detection of suspicious activities. For input surveillance footage, the overall goal of the proposed CWF based pre-processing method is to generate higher quality images that facilitate earlier and more accurate detection, thereby improving outcomes and efficiency. After preprocessing, the input data frames are extracted. Then the extracted frames are given to segmentation process.

3.3 Cell attention networks with U net (CAN-U-net) Based Segmentation

After preprocessing, segmentation is carried out by combining Cell Attention Networks with U-Net for Human Activity Recognition (HAR). The CAN-U-net is detail given below,

3.3.1 U-Net

U-Net is a type of convolutional neural network (CNN) architecture primarily used for image segmentation. U-Net based segmentation is a powerful technique used in various fields, including medical imaging and video surveillance. For Human Activity Recognition (HAR) and Suspicious Actions Detection in surveillance footage, U-Net can be particularly useful due to its ability to precisely segment and analyze video frames. U-Net consists of two main parts: an encoder (contracting path) and a decoder (expanding path). Each block in the path of contracting is made up of two consecutive 3 × 3 convolutions, a max-pooling layer, and a ReLU activation unit. There are multiple iterations of this arrangement. The second section of U-net, referred to as the expanded route, is where the innovation lies. Using 2 × 2 up-convolution, each stage expands the feature map. Next, the up sampled feature map is concatenated and cropped using a feature map from the matching layer in the contracting path. ReLU activation and two consecutive 3 × 3 convolutions come next. .. In order to create the segmented image and decrease the feature map to the necessary number of channels, a last step involves applying a further 1 × 1 convolution. Because the margins of the pixels contain the least contextual information, they must be cropped in order to remove them. As a result, a u-shaped network is created. More significantly, contextual information is propagated throughout the network, enabling it to partition items in a region by utilizing context from a broader overlapping region. The network's energy function is provided by equation (7):

$$F = \sum v(y) \log(p_{h(x)}(y)) \tag{7}$$

where, p_h represents a pixel wise SoftMax function applied over the final feature map. $v(y)$ denotes the activation. U-Net based segmentation can significantly enhance the ability to recognize human activities and detect suspicious actions in surveillance footage by precisely segmenting relevant regions and accurately interpreting actions within each frame.

3.3.2 Cell Attention Network(CAN)

Cell Attention Networks are designed to extend Graph Attention Networks (GAT) by incorporating multi-way relationships through edge-level attention mechanisms. When applied to Human Activity Recognition (HAR) for suspicious action detection in surveillance footage, can enhance the performance of segmentation models, particularly when integrated with U-Net architecture.

- **Cellular Lifting Map**

CAN begin by embedding input graphs into regular cell complexes using a skeleton-preserving cellular lifting map. This step ensures that the graph structure is preserved and facilitates the use of cell complexes for further processing.

- **Attention lift**

Attention lift learns edge features by applying a masked multi-head self-attention mechanism. It derives edge features from node features, allowing the network to capture complex relationships between connected nodes.

- **Cell Attention**

The core of CAN involves attention message-passing mechanisms at the edge level. As a result, the t-th message that passes round updates the edges embedding as equation (8):

$$\widetilde{k}_d^t = \varphi^t, (k_d^t, \underset{f \in M_j^t(d)}{\oplus} b_j^t(k_d^t, k_f^t)\psi_j^t(k_f^t), \underset{f \in M_v^t(d)}{\oplus} b_v^t(k_d^t, k_f^t)\psi_v^t(k_f^t) \in \mathrm{B}^{E^{t+1}}, \forall_d \in \Sigma^t, \quad (8)$$

where , $\oplus$ denotes the any permutation invariant operator. φ^t represents the learnable function. ψ_j^t and ψ_v^t are the learnable functions sharing the weights with b_j^t and b_v^t.

- **Edge Pooling**

After each message-passing round, edge pooling reduces complexity by retaining a subset of edges based on self-attention scores. This step simplifies the graph while preserving important features. Finally, a global readout aggregates the features from all layers, producing the final output which is then used for specific tasks such as classification or segmentation. Incorporating CAN with U-Net for HAR in surveillance footage can significantly enhance segmentation performance by combining the strengths of both architectures. U-Net's powerful feature extraction and segmentation capabilities are complemented by CAN's ability to model complex relationships and dependencies through edge-level attention, leading to improved detection of suspicious actions. After that, the segmented data is given to classification.

3.4 Convolutional Cross-Contextual Causal Dilated Neural Network(4CDNN) based classification

In the classification stage, the segmented data's are fed as input to the 4CDNN. Causal dilated convolutional neural networks with Cross-Contextual Attention based classification for Human Activity Recognition for Suspicious Actions Detection in Surveillance Footage can be broken down into several key components to explain the underlying concepts and their significance in the context of surveillance and suspicious actions detection.

3.4.1 Causal Dilated Convolutional Neural Network(CDCNN)

This is particularly useful when working with time series data such as ECG signals or surveillance videos as causal convolution shall guarantee that the output at depends only on the inputs at and previous times. This is desirable for applications such as HAR where future data is not available and hence the prediction has to be in real time. For the long sequences, it is required to manage more tokens for mechanical generation, and for this, disentangled causal convolution (DCC) is used. DCC adds a dilation factor the model is able to cover a larger time window without making the network too deep. This dilation improves the

receptive field exponentially; it helps develop the invariant by learning long range context from few layers. This is especially the case in HAR in surveillance videos because the activities last for different durations of time.

- **Improved Model Structure**

The suggested model for HAR contains multiple DCC blocks coupled with short cuts such as a dropout layer, weight normalize layer, activation function layer, a dilated causal convolution layer and a shortcut connection in every block. We also create a shortcut connection layer that links the input layer to the fully connected layer.

- **Dilated Causal Convolutional Layer**

In order to produce and capture long sequences, dilated causal convolution (DCC) is used. This dilation increases the receptive field exponentially, which captures the long-range dependencies with fewer layers. This is especially helpful to HAR in the surveillance video where activities take different intervals of time.

- **Weight Normalization Layer**

Weight Normalization Layer accelerates training by normalizing weights, leading to faster convergence. The neural network's functioning can be represented by equation (9):

$$x = \varphi(\mu \cdot y + a) \tag{9}$$

where, μ denotes the feature vector. a is a scalar parameter.

- **Activation Function Layer (ReLU)**

Introduces non-linearity and helps the model learn complex patterns. The activation function used in the model is ReLU .The ReLU activation function computation method is depicted in equation (10):

$$\mathrm{Re}\,LU(y) = \max(0, y) \tag{10}$$

where, y represents the input into the ReLU function. It can be any real number.

- **Dropout Layer**

Dropout is a regularization technique where random neurons are ignored during training, which prevents over fitting. By randomly setting the output of some neurons to zero, dropout ensures that the network does not rely on specific neurons, promoting generalization.

- **Shortcut Connection**

Neural networks with deeper network architectures typically exhibit the residual block structure. Through his research, he demonstrated that the learning effect will deteriorate when network depth increases beyond a certain point. The residual network adds shortcut connections to the deeper neural network, which facilitates network optimization. A residual block is made up of multiple network layers with a single, brief connection. The shortcut connection's computation expression is given equation (11):

$$o = (Y + E(Y)) \tag{11}$$

where, $E(Y)$ denotes the transformation applied to the input Y . E is the function of layers such as convolution layers, activation layers, activation function etc.. o denotes the output of the residual block. The combination of these components in the CDCN model allows for efficient and accurate classification of human activities, facilitating the detection of suspicious actions in real-time surveillance systems.

3.4.2 Cross-Contextual Attention

Cross-contextual attention-based classification leverages advanced attention mechanisms to improve the detection and classification of human activities, especially focusing on identifying suspicious actions in surveillance footage. This approach integrates information from multiple temporal contexts (e.g., different frames of a video) to enhance the accuracy and robustness of activity recognition. In the cross-attention module, features from the "before" frame (Y_1) serve as the Query, while features from the "after" frame (Y_2) is the key and value. Fixed sine spatial positional encoding is added to both Key and Value to maintain spatial relationships. Before computing the attention, the Query and Key features are normalized using $T2$ norm. This normalization helps in computing cosine similarity, enhancing the attention on bi-temporal relationships. The standard cross attention is given by equation (12):

$$X = O + soft\max(O \cdot W^{L}) \cdot Z \tag{12}$$

where, O denotes $\sum(Y_1, pos_1)$, W represents $\sum(Y_2, pos_2)$. Z denotes $\sum(Y_2, pos_2)$. Y shows the features of the output query. The cross-contextual attention-based classification method improves detection and classification of human activities in surveillance videos. It captures associations between features and context information, aiming for strong recognition. The method uses causal dilated convection and cross-contextual attention to detect suspicious behavior. BSTO is used to accurately distinguish between human actions and minimize 4CDNN error rate, processing time, complexity, and cost.

3.5 Optimization Using Boosted sooty tern Optimization (BSTO)

BSTO is an enhanced optimization algorithm optimizing sooty tern's foraging behavior to enhance 4CDNN learning parameters on energy consumption, convergence rate, and solution stability.

Step1: Initialization

Create an initial population of Boosted sooty tern Optimization solutions, each representing a set of 4CDNN hyper parameters.

Step 2: Generation of Random Variables

Generate at random the optimization variables of Boosted Sooty Tern optimization to attain the best solution.

Step 3: Evaluation of Fitness Function

The fitness for the t-th solution is defined by the objective function that is provided.

This fitness function's primary objective is to maximize classification accuracy while utilizing the fewest possible features. The fitness function equation is given by equation (13):

$$fitnessfunction = \delta + \eta \frac{|P|}{|E|} - D \tag{13}$$

where, P is the classification error rate . E indicates how many feature elements there are in the dataset. δ and η are the factors for weight and bias that indicate how important categorization quality . D denotes the classifier group column.

Step 4: Migration Behavior (Exploration) for improving accuracy

The process includes programming elements, such as sooty terns to act in an algorithm in

terms of collision avoidance and towards the best neighbor's direction. This includes computing new positions for the agents, directing them to the execution of the best performing neighbor positioning, and adjusting positions of agents regarding collision prevention and course. The agent further changes its position according to the selected ideal search agent. which is indicated by equation (14):

$$\vec{A}_{st} = \vec{B}_{st} + \vec{N}_{st} \tag{14}$$

where , st refers to sooty tern, $\vec{A}_{st}$ is used to bridge the gap between the ideal fitness search agent and the search agent.

Step 5: Attacking a behavior (Exploitation) for reducing error rate, processing time, computational complexity and cost

Sooty terns have the ability to alter their angle of attack and velocity by raising their wings above the ground. They spin into the air to attack their prey; this process is given by equation (15):

$$\vec{R}_{st}(x) = (\vec{C}_{st} a(a' + b' + x')).\vec{R}^{-}_{bst}(x) \tag{15}$$

where, $\vec{R}_{st}(x)$ maintains the best answer after updating the positions of other search agents.

Step 6: Termination

Once the best answers are obtained using equations (13-14), end the operation. Additionally, equation (15) yields the most accurate answer, and minimizes error rates, processing times, computing complexity, and cost. This iteration is remaining until the tentative criteria $j = j + 1$ is met. Lastly, the suggested 4CDNN-BSTO accurately recognizes the human activities for Suspicious Actions Detection in Surveillance Footage.

Hence, 4CDNN detailed and explained. It pre-processed data using CWF, Segmentation using CAN-U-net, Classification using 4CDNN optimization using BSTO method for activity of human recognition for Suspicious Actions Detection in Surveillance Footage demonstrating superior efficiency and accuracy. In the next section the results and discussions are discussed.

4. Result and Discussions

This section presents the experimental findings and comments of the proposed method performed to the Python platform.

4.1Dataset description

The KTH dataset is typical in that it includes 101 sequences for each of the six different activity types. Each sequence has almost 601 frames at a rate of 26 frames per second. This dataset is used to teach the model common behaviors, such as, jogging, running, boxing, hand waving, hand clapping and walking. In total, 7335 still photos have been assembled from various sources. Data that has been fully labeled by hand; 75% is utilized for training and 25% is for testing. Figure2 shows the KTH dataset Image of proposed method.

Figure2: KTH Dataset Image of proposed method

4.2 Performance metrics

The efficiency of the proposed method is evaluated based on various performance metrics such as Accuracy, Recall, and Precision, F- measure, Sensitivity, Specificity, and Error rate. The experimental outcomes are shown in the following table1:

Table 1: performance comparison of KTH Dataset

Methods	Accuracy (%)	Recall (%)	Precision (%)	F-measure (%)	Sensitivity (%)	Specificity (%)	Error Rate (%)
DDBN	84.93	32.12	42.00	26.00	79.66	84.33	1.0
EASAD	80.71	42.16	53.00	35.00	58.20	65.24	2.1
DCCGAN	72.67	40.20	44.00	37.00	68.18	76.56	1.5
DL	80.26	47.25	54.00	38.59	58.25	75.06	1.3
4CDNN (proposed)	99.8	48.70	66.00	42.00	80.68	90.00	0.1

Table 1 compares the performance of various methods on the KTH dataset across seven metrics: Accuracy, Recall, Precision, F-measure, Sensitivity, Specificity and Error rate. The proposed 4CDNN method outperforms others, achieving the highest values in Accuracy (99.8%), Precision (66.00%), F-measure (42.00%), Specificity (90.00%) and lowest error rate (0.1%). Figure 3 shows the training VS testing accuracy and loss,

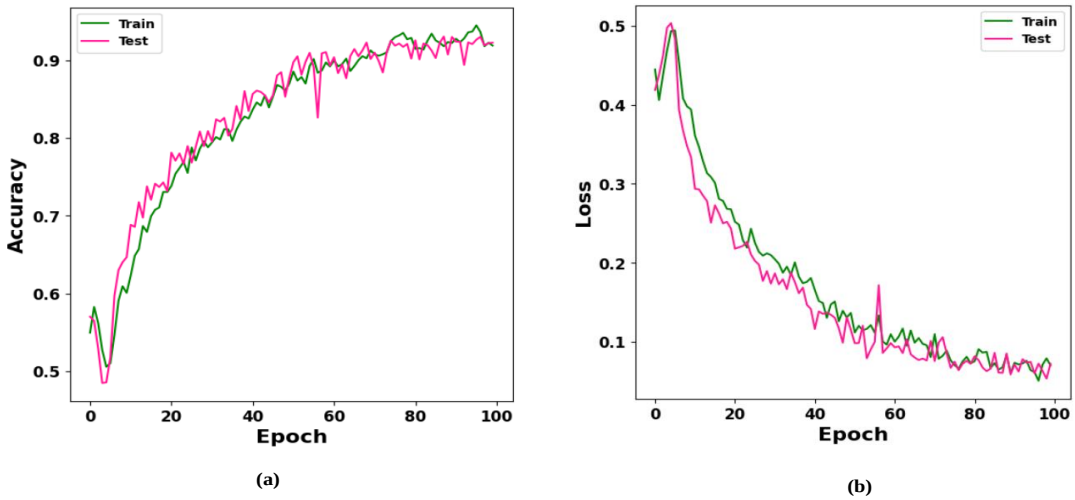

Figure 3: (a) Training VS Testing Accuracy (b) Training VS Testing Loss

Figure 3 illustrate the model's performance over 100 epochs. Figure (a) shows training and testing accuracy, both increasing and converging around 0.95. Figure (b) displays training and testing loss, both decreasing and converging below 0.1. The close alignment indicates good generalization, with minimal over fitting, demonstrating model robustness in detecting suspicious activities. The figure 4 shows the comparison of error rate with existing methods.

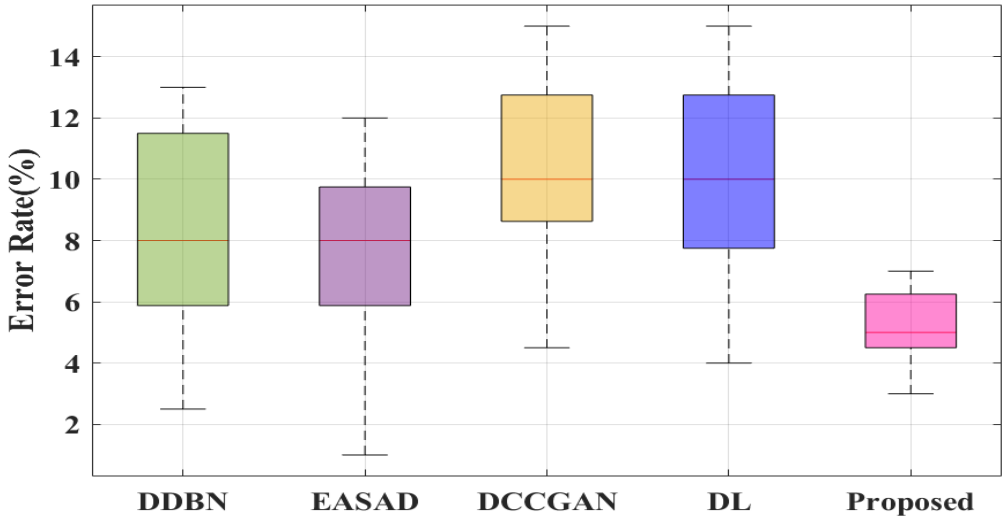

Figure 4: Comparison of Error rate with existing methods

Figure 4 compares the error rates of different methods: DDBN, EASAD, DCCGAN, DL, and the proposed method. The Proposed method exhibits the lowest error rate, with a median below 4%, indicating superior performance. The other methods have higher error rates, with medians around 8-12%, demonstrating less accuracy in comparison.

5. Conclusion

In this manuscript, 4CDNN-BSTO is successfully implemented. The input data is taken from KTH dataset. Then these data are pre-processed using CWF method. Following that, HAR is segmented using CAN-U-net method .Then the classification is done using convolutional Cross-Contextual Causal dilated neural network (4CDNN) for recognizing the human activities. The introduced system is executed in python. The efficiency of the proposed 4CDNN-BSTO is analyzed using KTH dataset and attains 99.8% accuracy and 0.1% error rate, compared with the existing methods. This indicates the approach's superior efficiency and potential for further development in the field. Future work will enhance model robustness and generalizability by expanding dataset, integrating real-time processing, and creating a user-friendly interface for detecting the HAR for the Suspicious Actions Detection in Surveillance Footage.

Reference

1. Parthasarathy, P., and S. Vivekanandan. "Detection of suspicious human activity based on CNN-DBNN algorithm for video surveillance applications." *2019 Innovations in Power and Advanced Computing Technologies (i-PACT)*. Vol. 1. IEEE, 2019.
2. Kumar, Manoj, Anoop Kumar Patel, and Mantosh Biswas. "Real-time detection of abnormal human activity using deep learning and temporal attention mechanism in video surveillance." *Multimedia Tools and Applications* 83.18 (2024): 55981-55997.
3. M. Preetha, Archana A B, K. Ragavan, T. Kalaichelvi, M. Venkatesan “A Preliminary Analysis by using FCGA for Developing Low Power Neural Network Controller Autonomous Mobile Robot Navigation”, International Journal of Intelligent Systems and Applications in Engineering (IJISAE), ISSN:2147-6799. Vol:12, issue 9s, Page No:39-42, 2024.
4. S. B, I. A. Karim Shaikh, P. Jagdish Patil, R. Sethumadhavan, M. Preetha and H. Patil, "Predictive Analysis of Employee Turnover in IT Using a Hybrid CRF-BiLSTM and CNN Model," 2023 International Conference on Sustainable Communication Networks and Application (ICSCNA), Theni, India, 2023, pp. 914-919, doi: 10.1109/ICSCNA58489.2023.10370093.
5. Kumar, Manoj, and Mantosh Biswas. "Abnormal human activity detection by convolutional recurrent neural network using fuzzy logic." *Multimedia Tools and Applications* 83.22 (2024): 61843-61859.
6. Srinivasan, S, Hema, D. D, Singaram, B, Praveena, D, Mohan, K. B. K, & Preetha, M. (2024), “Decision Support System based on Industry 5.0 in Artificial Intelligence”, International Journal of Intelligent Systems and Applications in Engineering (IJISAE), ISSN:2147-6799, Vol.12, Issue 15, page No-172-178.
7. A. Nithya, M. Raja, D. Latha, M. Preetha, M. Karthikeyan and G. S. Uthayakumar, "Artificial Intelligence on Mobile Multimedia Networks for Call Admission Control Systems," 2023 4th International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2023, pp. 1678-1683, doi: 10.1109/ICOSEC58147.2023.10275999.
8. Ahmed, Waqas, and Muhammad Haroon Yousaf. "A Deep Autoencoder-Based Approach for Suspicious Action Recognition in Surveillance Videos." *Arabian Journal for Science and Engineering* 49.3 (2024): 3517-3532.

9. K Sivakumar, M. Sughasiny, K.K.Thyagarajan, A. Karthikeyan & K. Sangeetha "A Comparative Analysis of GOA (Grasshopper Optimization Algorithm) Adversarial Deep Belief Neural Network for Renal Cell Carcinoma: Kidney Cancer Detection & Classification," International Journal of Intelligent Systems and Applications In Engineering, ISSN: 2147-6799, 2024, 12(9s), 43–48.
10. A. Jaya Mabel Rani, A Sumalatha, S Annie Angeline Preethi, R Indhu, M. Jayaprakash, M. Preetha, " Innovations in Robotic Technology: A Smar t Robotic Design for Effortless Wall Painting using Artificial Intelligence," 2024 International Conference on Automation and Computation (AUTOCOM), , doi: 10.1109/AUTOCOM60220.2024.10486126.
11. Waghchaware, Sheetal, and Radhika Joshi. "Machine learning and deep learning models for human activity recognition in security and surveillance: a review." *Knowledge and Information Systems* (2024): 1-32.
12. T.Mayavan, S. Sambath, A. Kadirvel, T.S.Frank Gladson, S. Senthilkumar, K Siva Kumar, "Artificial Neural Networks to the Analysis of AISI 304 steel sheets through limiting Drawing Ratio test" Journal of Electrical Systems, https://doi.org/10.52783/jes.3463 ISSN 1112-5209 2024, Vol: 20, 4s, 2282-2291
13. Shah, H., and M. S. Holia. "Multi-dimensional CNN Based Feature Extraction with Feature Fusion and SVM for Human Activity Recognition in Surveillance Videos." *Indian Journal of Science and Technology* 17.21 (2024): 2177-2198.
14. Nagalakshmi, P. "Human Activity Detection Events Through Human Eye Reflection using Bystander Analyzer." *Applied Artificial Intelligence* 38.1 (2024): 2321551.
15. Wani, Mohd Hanief, and Arman Rasool Faridi. "EASAD: efficient and accurate suspicious activity detection using deep learning model for IoT-based video surveillance." *International Journal of Information Technology* (2024): 1-13.
16. Kajendran, K., and J. Albert Mayan. "Recognition and detection of unusual activities in ATM using dual-channel capsule generative adversarial network." *Expert Systems with Applications* 247 (2024): 122987.
17. Kersten, Ivo PC, et al. "Detection of object throwing behavior in surveillance videos." *arXiv preprint arXiv:2403.06552* (2024).
18. Balaji Singaram, Lakshmi. B, Dr.M.Preetha, V.K. Ramya Bharathi, Dr.S.Muthumari lakshmi, Rakesh Kumar Giri "A Smart IoT-Based Fire Detection and Machine Learning Based Control System for Advancing Fire Safety",Nanotechnology Perceptions, ISSN 1660-6795 2024, Vol: 20, 5s, 229-244.
19. Karch, Barry K., and Russell C. Hardie. "Adaptive Wiener filter super-resolution of color filter array images." *Optics express* 21.16 (2013): 18820-18841.
20. Dr.M.Preetha, Balaji Singaram, Dr.I. Manju, B.Hemalatha, P. Bhuvaneswari "Machine Learning in Breast Cancer Treatment for Enhanced Outcomes with Regional Inductive Moderate Hyperthermia and Neoadjuvant Chemotherapy" Nanotechnology Perceptions, ISSN 1660-6795 2024, Vol: 20, 5s, 245-259
21. Siddique, Nahian, et al. "U-net and its variants for medical image segmentation: A review of theory and applications." *IEEE access* 9 (2021): 82031-82057.
22. Giusti, Lorenzo, et al. "Cell attention networks." *2023 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2023.
23. Balaji Singaram, M.S.Vinmathi, Dr.H.B.Michael Rajan, Jeyamohan H, T. Manikandan, "Data-Driven Estimation of Lithium-Ion Battery State-of-Health Prediction Approach Using Machine Learning Algorithm for Enhanced Battery Management Systems", Nanotechnology Perceptions, ISSN 1660-6795 2024, Vol: 20, 7s, 93-103.
24. Ma, Hao, et al. "An ECG signal classification method based on dilated causal convolution." *Computational and Mathematical Methods in Medicine* 2021.1 (2021): 6627939.