Video based Anomaly Detection Utilizing the Crow Search Algorithm-based Deep RNN

Laxmikant Malphedwar [#1], T. Rajesh [*2]
PhD Scholar, Saveetha School of Engineering,SIMTS Chennai, Tamil Nandu
Associate Professor, Saveetha School of Engineering,SIMTS Chennai, Tamil Nandu
laxmikant5656@gmail.com, t.rajesh61074@gmail.com

Article Info
Page Number: 10 - 23
Publication Issue:
Vol 71 No. 4 (2022)

Abstract
Numerous facets of our life have changed recently as a result of deep learning. The advancements in artificial intelligence enable computers to perform more of our regular chores. These days, there are more disruptive and provocative actions taking place than ever before. Security has thus been given priority consideration. CCTVs are being installed in more public locations, such as shopping malls, streets, banks, etc., to ensure people's safety. Because of this difficulty, a very accurate computerization of this system is now necessary. Since it would be very difficult for people to continuously monitor these security cameras. To determine whether the recorded actions are aberrant or suspicious, it demands workforces and their continual attention. Consequently, this shortcoming is driving a demand for highly accurate automation of this operation. The paper discusses the deep learning implementation technology that underlies the different crowd video analysis methods. For a variety of factors, including simplicity, performance, computational effectiveness, and high-quality interpretability, feature selection is crucial. Due to all the significance outlined above, an unique feature selection technique for anomaly detection that combines RNN and the crow search algorithm is proposed (CSA-RNN). In order to maximise the benefits of global search, the best attributes should be taken into account in each iteration. Additionally, it is necessary to identify which frames and sections of the recording include the odd activity in order to make a quicker determination of whether the unusual behaviour is suspicious or atypical. Its goal is to inevitably identify aggressive and violent behaviours in real time while removing variations from expected patterns. To recognize and categorize different levels of high movement in the frame, we wish to use a variety of Deep Learning model RNNs. From there, we may send out a danger detection alert, warning individuals of any ominous behaviour at a certain time.

Article History
***Article Received**: 25 March 2022*
***Revised**: 30 April 2022*
***Accepted**: 15 June 2022*
***Publication**: 19 August 2022*

Keywords: - Crow Search Algorithm, Deep Learning, RNN, Anomaly detection

1. Introduction

This template provides authors with most of the formatting specifications needed for preparing electronic A lot of research has been done recently in the area of classifying human behaviour for automated monitoring. The cataloguing or classification of observed communicative occurrences using an algorithm is known as behaviour classification. This study endeavor has been motivated by a rise in security and safety concerns as well as an abundance of surveillance data compared to the amount of processing capacity available. A subset of behaviour classification issues that is simplified tasks is anomaly detection in computerized scrutiny. CCTV in a setting collects data that represents the behaviour of surveillance targets as part of the automated monitoring process, with some behaviours being presumed to be aberrant. Then a feature extraction procedure is used to the raw videos. The collected attributes go into a modelling system that uses a learning technique to assess if the observed behaviour is normal or abnormal. can be used offline or in real time; in the latter instance, feature extraction and modelling are performed to documented data. This evaluation will largely concentrate on instantaneous solutions; for more detail relevant to the offline scenario, consult Saykol's review [1].

By focusing on a specific area of the data and ignoring the vast volumes of meaningless statistics, automated anomaly detection is incredibly helpful in decreasing the quantity of data that can only be addressed manually. When used in conjunction with an object identification system, video surveillance is a crucial tool for examining city crossings, road networks, and pedestrian movement. Additionally, aberrant object identification enables better monitoring and trailing of the object of interest, allowing for the timely implementation of protective or cautious changes to the event that is unfolding. However, there is a lot of opportunity for perception when it comes to the problem of anomaly detection, and research efforts vary not just in terms of technique but also in terms of how the tricky, underlying conventions, and desired results are understood.

Studies already conducted on these issues use a variety of classical techniques for detection [2- 4]. These studies have provided persuasive evidence in straightforward scenarios, but their capacity to address compelling conditions in complex situations is rather constrained. Most of the research that have already been done used a single camera view and trained their models on tiny datasets with specified limitations for identifying specific items. Using an established method on a device with limited resources raises the computational cost, which results in subpar object detection performance. The majority of video surveillance panels include many cameras, thus an object detection system that operates simultaneously is crucial.

Additionally, it may be said that one of the most significant and popular study subjects is input optimization. [5]–[7]. It is a part of every aspect's fundamental operations and is present in practically every industry, including engineering, science, energy, computer science, etc. [8]–[10]. Optimization has become a significant difficulty due to the real-world scientific and engineering problems' rising complexity. A novel system called Crow Search Algorithm (CSA) mimics how crows store their food and get it out when they need it. Since its introduction, CSA has found extensive usage in a variety of optimization problems, including those in chemical engineering, medicine, power and energy, feature selection, and image processing. Consequently, we optimize using this technique in our study.

The suggested model has the capability to extricate between the normal and abnormal classes. Multiclass categorization, which has the two dominant classifications of normal and abnormal, is used. A frame wherein the aberrant object activities take place belongs to the abnormal class, which is divided into two separate abnormal subclasses depending on these activities' features. The important class that corresponds to an allowed subclass activity is specified by the subclass. Similar to this, the normal class, which is further divided into two normal subclasses, is a frame that contains normal object actions.

The following is a list of the study's key contributions:

1. To start, a feature extraction method known as CSA has been described. It allows us to choose a small subset of the important aspects of the original components while removing any extraneous or superfluous elements.

2. The subsequent network is utilised as a recurrent neural network to infer sense since the series of behaviours shown over a predetermined amount of time. Using this process, the components of movies will be classified as either safe or harmful.

3. Rapid, precise, and effective identification of abnormal events from massive amounts of video data without human involvement or oversight.

2. Problem Formulation

An algorithm that draws inspiration from nature called the 'crow search algorithm' (CSA) [11] can be utilised to offer the best answers to various issues. Askarzadeh introduced CSA for the first time in 2016 by incorporating the clever behaviours of crows, such as thieving or concealing food, into a mathematical model. Because they track further crows to snip their food, utilise their memory to recall faces, are self-aware, and live in groups, crows are thought to be among the most intellectual animals [30, 31]. Crows constantly shift the position of their concealed food in order to protect it. Crows utilise probabilities to detect potential thievery.

Two major categories are used to discuss the literature on anomaly detection techniques: Deep feature-based approaches and conventional handmade feature-based methods for identifying abnormal events. Anomaly detection used to be very reliant on subordinate, manually created feature-based approaches. These techniques principally rely on three phases: (1) feature extraction, which involves extracting low-level habits from the training set; (2) feature learning, which is illustrious by the dissemination of encoding predictability or normal events; and (3) outlier detection, which involves identifying detached gatherings or outliers as inconsistent measures. For instance, Zhang et al. [12] used spatiotemporal characteristics to characterised frequent encounters using the Markov random field. Similar to this, a social interaction model [13] where assistance services were calculated and normal and aberrant behaviours were identified using optical flow is created.

Recent years have seen an increase in the importance of automated video surveillance systems as study topics in the realm of public safety. There has been a lot of exertion testified on tracking and recognizing an object's movement. Another significant factor in the decrease in workload and rise in surveillance effectiveness is artificial intelligence. Anomaly recognition is a difficult and well-established problem in computer vision [14][15]. In a conference paper [16], a method for using sensor systems that can notify on the existence of any apprehensive behaviour is suggested. Events are accounted for by sensors, but no information is given about them. In order to take preventative measures, information regarding the threat and its root cause may be rapidly and correctly gathered by analysing the recorded footage. Therefore, it's possible that using CCTV camera and sensor systems alone or in combination won't be enough to identify undesirable occurrences in real time. As a result, they created a system that uses a camera and sensor networks to quickly identify threats in a variety of lighting situations. An intelligent, quantifiable, clever, and unwavering algorithm helped create the system. To evaluate the human shape and record the human motion, body-based detection and multiple edge detection techniques were used, along with background estimation and edge detection. Recently, Mohammadi et al. [17] suggested using a behaviour heuristic-based technique to cope with the categorization of violent and calm recordings. In addition to identifying violent and typical patterns, other writers suggested using tracking to identify anomalies and characterised strangeness as a departure from that typical movement. An essay by Bharath Raj [18] describes how to use object detection as part of a Deep Learning-based surveillance architecture. Deep learning and CNN have been used (19) to spot drive and categorize objects in videos.

According to the literature analysis conducted thus far, the majority of studies have developed methods for teaching the distribution of common motions through practice utilizing recordings that are already accessible. Additionally, during testing, certain patterns that cause a significant restoration error are deemed abnormal. Deep learning has shown to be the most effective method for classifying images, making it appropriate for classifying video content. Encoders built on deep learning have been used to inevitably train the archetypal for typical behaviour and engage rebuilding loss to identify anomalies. So, in order to achieve efficient results, we incorporate CSA-based RNN into our suggested system.

3. Projected system

3.1 Crow Search Algorithm (CSA):

An innovative meta-heuristic optimization practice called the "crow search algorithm" (CSA) is based on how crows save their extra food in secret locations and then find it when they need it. Crows have enormous brains compared to their bodies, which allows them to recall the appearances of other crows and caution one another when a dangerous crow slants. They also possess advanced communication skills and long-term memory, remembering where food is buried for quite a lot of months. Consequently, the key ideas of crows are that they watch where other birds get their food and steal when there are no other birds present. Additionally, if a crow has stolen anything formerly, it will take superfluous steps like fluctuating hiding spots to avert being a victim again.

Assume that N is the quantity of crows and that d is a multidimensional environment. Crows i's location at iteration iter is determined by:

$$x^{i,iter} = \left[x_1^{i,iter}, x_1^{i,iter}, \dots, x_1^{i,iter}\right] \qquad (1)$$

where i = 1, 2, · · · , N and iter = 1, 2, · · · , itermax. itermax is the maximum number of iterations.

Crows travel throughout the surroundings looking for better food sources (walloping places). Each species has a reminiscence that allows it to remember where its food hiding spots are. Following Crow j to get near to the location of the food's hiding spot and hunting it is one of the crow's primary hobbies. Consequently, CSA may experience two significant cases:

Crow j is unaware that i is behind it. Crow i will thus approach j's walloping location and take the following new position:

Case 1: Crow j does not recognize that i is subsequent to it. Hence, i will approach the hiding location of j and take a new site with respect to given equation:

$$x^{i,iter+1} = x^{i,iter} + r_i . fl^{i,iter} . (m^{j,iter} - x^{i,iter}) \qquad (2)$$

where ri is a random number, $fl^{i,iter}$ point to the flying extent of i (see Fig 1), and $m^{j,iter}$ epitomizes the best stayed site by j.

Case 2: Crow j perceives that it is trailed by i. Crow j will thus alter the flying permit to deceive factions by relocating to a different point in the search space in order to prevent its cache from ever being stolen.

Therefore, the location can be articulated as:

$$x^{i,iter+1} = \begin{cases} x^{i,iter} + r_i . fl^{i,iter} . (m^{j,iter} - x^{i,iter}), & r_j \geq AP^{j,iter} \\ a\ random\ position & Otherwise \end{cases} \qquad (3)$$

where $AP^{j,iter}$ identifies each crow's likelihood of being conscious.

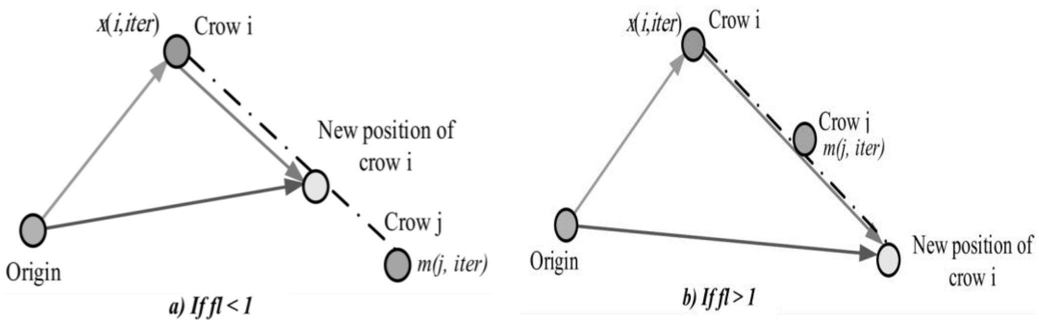

Fig 1: The effects of the crow flight length on the development of the search

To address the problem of detecting anomalous behaviour, we suggest the CSA technique. Finding the best combination of control variables to minimise the optimization problem while meeting all power system requirements is the major goal. The following stages are used to implement the CSA for the problem of anomalous behaviour detection:

Step 1: Setting up the algorithm's parameters and restrictions

Set the amount of crows (N), the number of iterations ($iter_{max}$), the flying distance (fl), and the awareness probability (AP) to their initial values. Identify the restrictions and decision factors.

Step 2: Set up the crows' location and memory

N crows should be generated at random. Each crow offers a suitable response to the issue. Crows are said to have concealed their food in the early placements since they are thought to be inexperienced at first.

Step 3: Assess fitness performance

By modifying the standards of the regulator variable in the objective function, the quality of each crow's location is calculated.

Step 4: Create a new position.

The following is how The Crow *i* creates a new position: To locate the location of the concealed food, it randomly chooses one of the other crows and follows it. Equation (3) therefore provides the new location of crow *i*. Each crow goes through the same process.

Step 5: Check the viability of new jobs

Each crow's new location is tested for viability. A crow updates its position if its new location is accurate. Otherwise, it does not change to the newly generated position and instead remains in its present place.

Step 6: Assess the fitness requirements for new roles

The fitness function value of the new site is calculated for each crow.

Step 7: Upgrade memory

Crows apprise their collective memory as trails if their estimate is superior to the remembered fitness function value:

$$m^{i,iter+1} = \begin{cases} x^{i,iter+1} & f(x^{i,iter+1}) \text{ is better than } f(\, m^{i,iter}), \\ m^{i,iter} & Otherwise \end{cases} \quad (4)$$

where *f*(.) indicates the objective function value.

Step 8: Verify the break criteria

Up till $iter_{max}$ is attained, repeat Steps 4–7. The best memory place that corresponds to the best objective function value is the first answer to the anomaly detection issue.

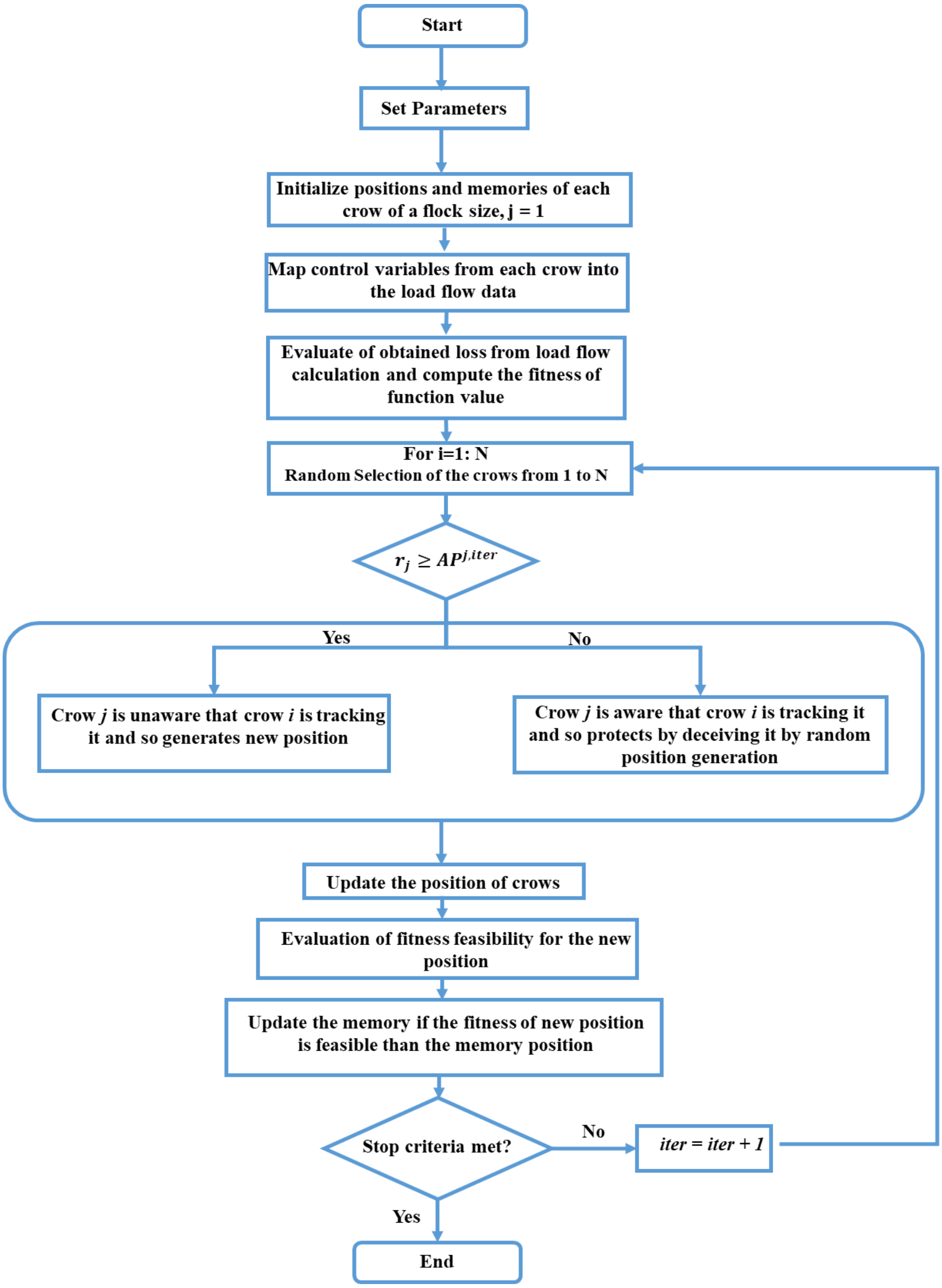

Fig 2: Flow diagram of Crow search algorithm

3.2. Detection Approach Using Recurrent Neural Networks.

After choosing the ideal position, categorization of the input frames and detection are crucial. Traditional neural networks cannot handle many contextual problems because neurons in the same layer do not communicate with one another. RNN is typically thought of as a neural network that operates according to time sequence and offers special advantages for processing time series jobs due to the particularity of its network. By preserving some knowledge about the previously processed content, it reaches a certain memory ability and replicates how individuals would read an article in order to better understand upcoming content. Figure depicts the recurrent neural network's structural diagram.

After the determining best position, the classification of frames from the input and detection plays a major role. Traditional neural networks cannot handle many contextual problems because neurons in the same layer do not communicate with one another. RNN is typically thought of as a neural network that operates according to time sequence and offers special advantages for processing time series jobs due to the particularity of its network. By preserving some knowledge about the previously processed content, it reaches a certain memory ability and replicates how individuals would read an article in order to better understand upcoming content. Fig. 3 shows the recurrent neural network's structural diagram.

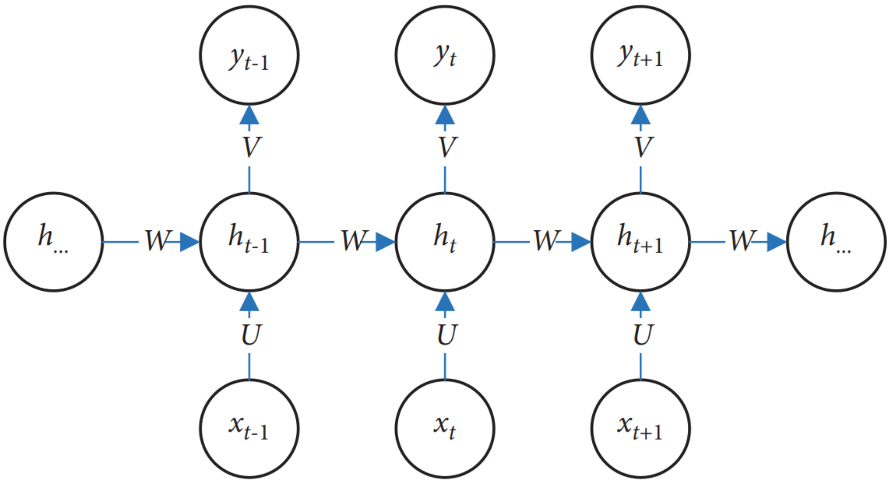

Fig 3 RNN structure

Figure 3 shows the input as x, the hidden state as h, the prediction result as y, and the weight matrix as W, U, and V. The behaviour and position of the network must be examined in terms of time as RNN is a time series framework. The network's input sequence xt and the network state ht−1 at the previous time define the neuron state ht at time t. The neuron state ht at this moment may be determined as follows:

$$h_t = f(Ux_{t-1} + Wh_{t-1} + b_h) \qquad (5)$$

where f is instigation function and bh is a bias term.

The outcome of the network state at time t + 1 is the neuron state ht at time t, which is also utilised as the input of the system state at that time. Nevertheless, as a result, ht can indeed be output directly. It must be multiplied by a coefficient V, followed by the addition of the offset ob, and normalisation is necessary. The following is the mathematical computation formula:

$$y_t = act(Vh_t + b_y) \tag{6}$$

where act is activation function and by is a bias term.

The model's parameters are shared by the RNN at various points, which decreases the number of parameters that must be learned but causes extremely unstable model parameters while updating. However, nearby are concerns with gradient expansion or gradient vanishing, and as a result, only short-term memory is accessible. RNN theoretically has the potential to cope with long-term reliance matters.

One of the popular ones is the LSTM network. An RNN with a unique structure is the LSTM. By include gating units in the model, it decides whether input information is remembered or forgotten, thereby resolving the issue of lengthy sequence dependency. In plain English, LSTM performs better than regular RNN while handling longer sequence problems.

Three gating units and a cell unit make up an LSTM model. The relevant content data is stored in the cell unit. The stuff transmitted from the previous moment is filtered by the Forgotten Gate, which keeps pertinent information and discards irrelevant data. The input gate keeps the network input only when it makes sense to do so, thereby removing unnecessary data. The output gate outputs the cell's current state on a selective basis. The following is the calculating formula:

$$I_t = \sigma(W_I x_t + A_I h_{t-1} + b_I) \tag{7}$$

$$P_t = \sigma(W_P x_t + A_P h_{t-1} + b_P) \tag{8}$$

$$O_t = \sigma(W_O x_t + A_O h_{t-1} + b_O) \tag{9}$$

$$c_t = W_t c_{t-1} + I_t \tilde{c}_t \tag{10}$$

$$\tilde{c}_t = tanh(W_C x_t + W_C h_{t-1} + b_C) \tag{11}$$

$$h_t = O_t \tanh(c_t) \tag{12}$$

Where W_I, W_P, W_O, and W_C represents input weight vectors, while A_I, A_P, A_O, and A_C represent upper output weight vectors. Then b represents bias vectors; σ= sigmoid function.

Figure 4 displays the fundamental architecture for recurrent neural network-based anomaly detection. The main idea is to analyse the intrusion data by obtaining correlation data between attributes, create a detection model, and utilise the time series handling capabilities of the recurrent neural network. The basic procedure is as follows: (1) developing and training a recurrent neural network model; (2) applying the learnt model to classify and predict unknown data; and (3) employing preprocessing to convert the initial dataset into data in a standard format.

The CSA-RNN approach is comprised of preprocessing, CSA feature extraction, frame classification, localization, and detection, as illustrated in Figure 4. The ensuing elements combine in a certain order to discover the required abnormality.

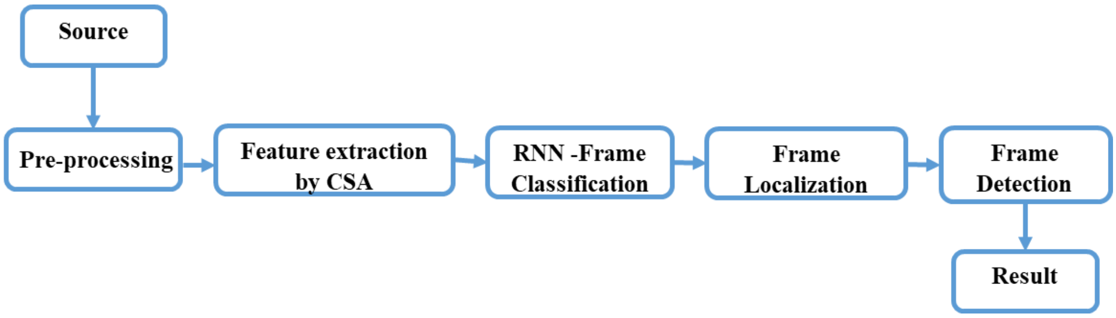

Fig 4: Proposed Model of CSA-RNN

The initial step in this method is to extract frames from the CCTV recordings that were obtained. After a predetermined and brief period of time, the process extracts the frame (say 1 sec). This extracted frame underwent preprocessing in order to properly adapt the scaled picture to the format required by the model. The concatenated group of high-level feature maps created in the preceding phase serves as the neural network's input.

In essence, frame detection is utilised to precisely locate and measure an item in an image, which is crucial for frame classification [20]. Frame localization establishes an object's location and size [21]. The learning algorithm is carried out in the following phases after the architecture has been defined: In the neural network learning process, the weights of each defined layer's parameters are essentially calculated incrementally. The objective is to obtain the most accurate forecast parameters.

4. Experimental Results:

4.1. Dataset and Evaluation Metric.

Avenue Dataset, which is utilised in the experiment, includes 16 training and 21 testing video clips. We made advantage of the neural network training platform TensorFlow [22].

Two neurons make up the output layer of the RNN model, which is used to divide the complete dataset into two sorts, threat and safe. The videos utilised are not edited and include a number of undesirable scenes. There are 940 chunks of unshuffled frames, extracted in 30 frame batches at 1 second intervals. Preprocessing and feature extraction help to lessen them.

There is a discussion of the evaluation and comparison of the current research with the suggested technique.

4.2. Quantitative Results Analysis

We employed a confusion matrix, a table that details how well the classification model performed. Accuracy, precision, recall, and F1 measure are the metric parameters used for assessment and comparison [23], and they may be characterised as follows:

Accuracy: It is defined as the number of variables in the dataset that have been effectively categorised.

$$\text{Accuracy} = (TP + TN)/ (TP + TN + FP + FN)$$

Precision: It is defined as the proportion of correctly categorised things to ones that are incorrectly categorised.

$$\text{Precision} = TP / (TP + FP)$$

Recall: Recall is whether any of the positive instances the classifier confidently predicted across all of the positive examples in the data. Sensitivity is a term that is occasionally used to describe it.

$$\text{Recall} = TP / (TP + FN)$$

F1-score: It's defined as the number of objects that have been categorised erroneously.

$$\text{F1-score} = 2 * TP/ (2 * TP + FP + FN)$$

When compared to other approaches of existing methodologies like CNN, F-CNN and proposed method. The following Table 1 provides a clear picture of reduced complexity and the best accuracy rate.

Table 1: Performance Metrics of Proposed system

Methods	Accuracy	Precision	Recall	F1 Score
CNN	89.7	84.9	81.6	82.5
F-CNN	92.6	87.2	84.6	87.4
RNN	96.8	96.5	94.5	96.1

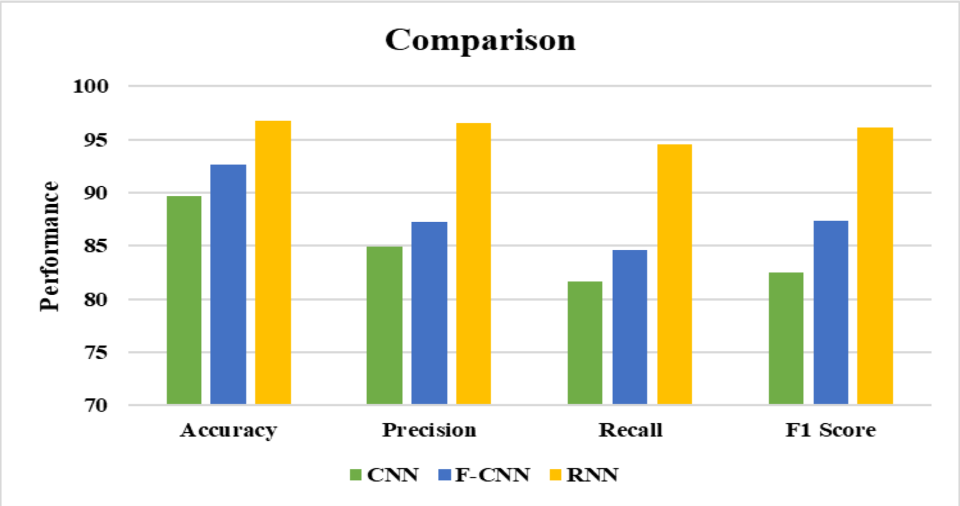

Fig 5: Comparison graph

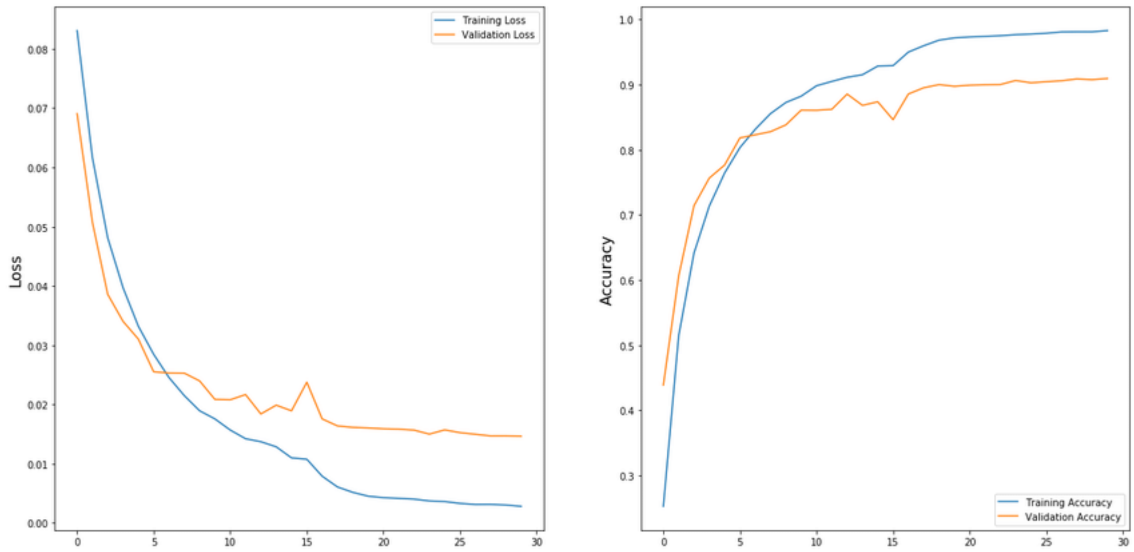

Fig 6: Training and Validation Loss and accuracy

The graph above demonstrates how the suggested approach increased testing accuracy to a significant degree and how the loss function is also meeting more quickly, as illustrated in Fig. 6. *a* and *b*, respectively As a result, the model's overall performance has been improved by all of the improvements that have so far been included.

5. Conclusion

This study offers a method for identifying outliers in actual CCTV footage. It might not be possible to identify anomalies in these recordings using only the standard data. The model's accuracy has therefore been maximized by taking into account both normal and anomalous films in order to manage the intricacy of these realistic abnormalities. Additionally, a universal typical of anomaly identification has been learnt using two different neural networks with a weakly categorized dataset in order to avoid the labor-intensive chronological labelling of anomalous regions in training recordings. The experimental findings collected during the investigation lead to the conclusion that our proposed anomaly detection strategy outperforms the previously employed approaches substantially.

Hence, the overall accuracy of the model is 96.8% with reduced overfitting.

6. Future Scope

The outcomes of deep learning replicas for identifying apprehensive behaviour are presented in this article. The outcome of the recognition produced by these models demonstrates the importance of our dataset and opens up possibilities for more research.

- Future expansion of this project's features is possible:
- The real-time model execution in an anomaly recognition system is difficult. In the future, a more cost-effective and efficient approach can be used to address issue. The model may also be

enhanced to find prospective threats and notify the relevant authorities in front of them, improving public safety.

- The model might be expanded to incorporate a wider range of abnormalities.

Reference:

[1]. E. Saykol, U. Gudukbay, and O. Ulusoy, "Scenario-based query processing for video-surveillance archives," *Eng. Appl. Artif. Intell.*, vol. 23, no. 3, pp. 331–345, 2010.

[2]. Gesick, R.; Saritac, C.; Hung, C.C. Automatic image analysis process for the detection of concealed weapons. In Proceedings of the 5th Annual Workshop on Cyber Security and Information Intelligence Research: Cyber Security and Information Intelligence Challenges and Strategies, Oak Ridge, TN, USA, 13–15 April 2009; pp. 1–4.

[3]. Flitton, G.; Breckon, T.P.; Megherbi, N. A comparison of 3D interest point descriptors with application to airport baggage object. detection in complex C.T. imagery. *Pattern Recognit.* **2013**, *46*, 2420–2436.

[4]. Zhang, X.; Wang, L.; Su, Y. Visual place recognition: A survey from deep learning perspective. *Pattern Recognit.* **2021**, *113*, 107760.

[5]. X. Zhao, D. Li, B. Yang, C. Ma, Y. Zhu, and H. Chen, ''Feature selection based on improved ant colony optimization for online detection of foreign fiber in cotton,'' *Appl. Soft Comput.*, vol. 24, pp. 585–596, Nov. 2014.

[6]. M. Wang and H. Chen, ''Chaotic multi-swarm whale optimizer boosted support vector machine for medical diagnosis,'' *Appl. Soft Comput.*, vol. 88, Mar. 2020, Art. no. 105946.

[7]. X. Zhao, X. Zhang, Z. Cai, X. Tian, X. Wang, Y. Huang, H. Chen, and L. Hu, ''Chaos enhanced grey wolf optimization wrapped ELM for diagnosis of paraquat-poisoned patients,'' *Comput. Biol. Chem.*, vol. 78, pp. 481–490, Feb. 2019.

[8]. Y. Sun, B. Xue, M. Zhang, G. G. Yen, and J. Lv, ''Automatically designing CNN architectures using the genetic algorithm for image classification,'' *IEEE Trans. Cybern.*, vol. 50, no. 9, pp. 3840–3854, Sep. 2020.

[9]. H. Chen, Y. Xu, M. Wang, and X. Zhao, ''A balanced whale optimization algorithm for constrained engineering design problems,'' *Appl. Math. Model.*, vol. 71, pp. 45–59, Jul. 2019

[10]. H. Chen, M. Wang, and X. Zhao, ''A multi-strategy enhanced sine cosine algorithm for global optimization and constrained practical engineering problems,'' *Appl. Math. Comput.*, vol. 369, Mar. 2020, Art. no. 124872, doi: 10.1016/j.amc.2019.124872

[11]. Askarzadeh, A. (2016). A novel metaheuristic method for solving constrained engineering optimization problem: Crow search algorithm. *Computers & Structures, 169*, 1–12.

[12]. Zhang, D.; Gatica-Perez, D.; Bengio, S.; McCowan, I. Semi-supervised adapted hmms for unusual event detection. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–25 June 2005; Volume 1, pp. 611–618.

[13]. Mehran, R.; Oyama, A.; Shah, M. Abnormal crowd behavior detection using social force model. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 935–942.

[14]. Waqas Sultani, Chen Chen, Mubarak Shah. "Real-World Anomaly Detection in Surveillance Videos", 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018.

[15]. Junseok Kwon. "Rare-Event Detection by QuasiWang –Landau Monte Carlo Sampling with Approximate Bayesian Computation", Journal of Mathematical Imaging and Vision, 2019.

[16]. Gaurav Kumar Singh, Vipin Shukla, Pratik Shah. Automatic Alert of Security Threat through Video Surveillance System. In 54th Institute of Nuclear Material and Management Annual Meeting, 2013.

[17]. S. Mohammadi, A. Perina, H. Kiani, and M. Vittorio. Angry crowds: Detecting violent events in videos. In ECCV, 2016.

[18]. Bharath Raj, How to Automate Surveillance Easily with Deep Learning, https://medium.com/nanonets/how-to-automate-surveillance-easily-with-deeplearning- eb4fa0cd68d, 2018.

[19]. A Gentle Introduction to Object Recognition with Deep Learning, https://machinelearningmastery.com/object-recognition-with-deeplearning/, 2019.
[20]. Felzenszwalb, P.F.; Girshick, R.B.; McAllester, D.; Ramanan, D. Object detection with discriminatively trained part-based models. IEEE Trans. Pattern Anal. Mach. Intell. 2009, 32, 1627–1645.
[21]. Géron, A. Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow: Concepts, Tools, and Techniques to Build Intelligent Systems; O'Reilly Media: Sebastopol, CA, USA, 2019.
[22]. Baskurt, K.B.; Samet, R. Video synopsis: A survey. Comput. Vis. Image Underst. 2019, 181, 26–38
[23]. Verma, G.K.; Dhillon, A. A handheld gun detection using faster r-cnn deep learning. In Proceedings of the 7th International Conference on Computer and Communication Technology, Allahabad, India, 24–26 November 2017; pp. 84–88.
[24]. https://en.wikipedia.org/wiki/Corvus28genus%29;https://doi.org/10.1016/j.neucom. 2016.12.038.
[25]. Rincon, P. Science/nature|crows and jays top bird IQ scale, BBC News