

An NFMF-DBiLSTM model for human anomaly detection system in surveillance videos

Sanjeevkumar Angadi¹, Chellapilla V. K. N. S. N. Moorthy², Mukesh Kumar Tripathi³,
Bhagyashree Ashok Tingare⁴, Sandeep Uddhavrao Kadam⁵, Kapil Misal⁶

¹Department of Computer Science and Engineering, Nutan College of Engineering and Research, Pune, India

²Department of Mechanical Engineering, Vasavi College of Engineering, Hyderabad, India

³Department of Computer Science and Engineering, Vardhaman College of Engineering, Hyderabad, India

⁴Department of AI and DS, D. Y. Patil College of Engineering, Pune, India

⁵Department of Computer Engineering, Keystone School of Engineering, Pune, India

⁶Department of Master of Computer Application, Trinity Academy of Engineering, Pune, India

Article Info

Article history:

Received Jan 8, 2024

Revised Jun 1, 2024

Accepted Jun 25, 2024

Keywords:

Machine learning

NFMF-DBiLSTM

PEOO algorithm

SA-KM clustering algorithm

ABSTRACT

In response to the increasing demand for an intelligent system to avoid abnormal events, many models for detecting and locating anomalous behaviors in surveillance videos have been proposed. Nevertheless, significant flaws of inadequate discriminating ability are present in the majority of these models. A novel newton form and monotonic function based deep bidirectional long short-term memory (NFMF-DBiLSTM) human anomaly recognition system was discussed in this paper to tackle those issues. Initially, videos are transformed into frames; after that, the duplicate frames are removed, and by utilizing the shannon entropy centered contrast limited adaptive histogram equalization (SE-CLAHE) algorithm, the contrast has been elevated. By using the probabilistic matrix factorization kernel density estimation (PMF-KDE) technique, the background is subtracted after estimating only the motion of the object. After this, the silhouette function is performed utilizing the dirac depth silhouette function (DDSF). In addition, clustering is done by sorting and average-based K-means (SA-KM). The features are extracted from the suspected human and are then chosen by utilizing Poisson Eurasian oystercatcher optimization (PEOO). For classifying normal or anomaly, the selected features are subjected directly into the NFMF-DBiLSTM. When contrasted with the prevailing methodologies, the proposed model is found to be more efficient.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Chellapilla V. K. N. S. N. Moorthy

Department of Mechanical Engineering, Vasavi College of Engineering

Hyderabad, India

Email: krishna.turbo@gmail.com

1. INTRODUCTION

The utilization of visual surveillance along with cameras for security applications is becoming more familiar with the rise in crime around the world; in addition to that, it has become a recognized part of modern life. In places that require to be secured, closed-circuit television (CCTV) is utilized for video surveillance [1]-[3]. Automatically and effectively determining possible anomalies or objects of interest from a large number of surveillance videos has become a challenging task [4]. Owing to the progress of research in recent years, a broad range of modern approaches for anomaly detection (AD) from surveillance have been developed [5], [6]. It can play a vital part in detecting/predicting accidents, congestion, and other anomalies besides gathering statistical details about the road traffic's status. Thus, various computer vision-centric

studies and challenges, namely traffic monitoring, activity recognition, emergency management, human behavior analysis, event detection, and so forth, have been conducted [7].

AD from surveillance scenes is a subdomain of behavior understanding. The detection of unexpected, unusual, irregular, or unpredictable events or items, which are not considered to be generally occurring events or regular items in a pattern or items available in a dataset and hence differ from prevailing patterns, is known as AD. An anomaly is a pattern that differs from the standard patterns' set. Consequently, anomalies are dependent on the phenomenon of interest [5]. In today's public security, safety, sports analysis, group activity monitoring, and visual surveillance, automatic AD is critical [8]. Automated surveillance systems can greatly help in making suitable decisions for safety along with emergency control since they predict uncommon and complicated situations in a congested environment [9]. As an outcome, to detect and regulate crowds for public security, safety, along with statistical intentions, surveillance techniques are needed in difficult and congested environments, namely political rallies, busy streets, airports, shopping malls, public celebrations, and train stations [10].

Grounded on behavior representation, AD models are generated in the prevailing works. By extracting some important features, namely [11], [12] optical flow, spatiotemporal volume, shape, trajectory, et cetera, the detection can be performed [13]. By executing the approaches, namely principal component analysis (PCA), the support vector machine (SVM), the k-nearest neighbor (k-NN) algorithm, or various sorts of correlation analysis, behavioral features are further used. The objective of detecting anomalies utilizing static and time-variant systems was shared by all frameworks. Additional approaches, namely sliding windows (SWs), are employed in conjunction with the abovementioned techniques to detect dynamic and time-variant anomalies. As an outcome, systems that sufficiently capture time-variant model dynamics are not available in the approaches and hence, anomalous context cannot be characterized. In recent years, various research has been conducted on machine learning (ML), namely long-short term memory (LSTM), recurrent neural network (RNN), convolutional neural network (CNN), and so on, to tackle these problems. Grounded on characteristics of time series, encompassing univariate and multivariate time series data approaches, the prevailing detection approaches were classified by the survey. Nevertheless, since the occlusion issue is not considered, these approaches are not rendering the desired result.

Tripathi *et al.* [14], [15] explained real-time crowd AD utilizing a convolutional long short-term memory (Conv-LSTM) network. A deep learning (DL)-centric approach was utilized to predict violent activities and aid stakeholders in displaying such activities in real time. Conv-LSTM was utilized to capture the frame along with detect violent actions. Better accuracy was obtained by the suggested system at a faster rate. Nevertheless, accuracy was still lacking owing to the complexity of individual or group activity classification. Shin *et al.* [16] proposed an online AD technique for surveillance videos utilizing transfer learning as well as continual learning. The developed algorithm utilized the feature extraction (FE) power of neural network-centered methodologies and statistical detection methods. Simulation outcomes considerably gave pre-eminent accuracy for the built system. Nevertheless, it was still challenging to learn to detect abnormalities promptly. Ullah *et al.* [17], [18] established an anomalous event detection technique by utilizing weakly labeled training videos. For FE, a deep residual learning framework I3D-residual network (Resnet)-50 was utilized. Significant improvement in the outcomes regarding both accuracy and recall was acquired by the introduced system for video AD.

2. PROPOSED METHODOLOGY OF HUMAN ANOMALY DETECTION SYSTE

By utilizing a novel NFMF-DBiLSTM model, detecting human anomalies in video surveillance is proposed in this paper. Videos are converted into frames in this model. Some major steps like contrast enhancement, background subtraction, and clustering are executed after the conversion. Later, the features are extracted and selected. Lastly, for classification as normal or anomaly, the selected features are fed into the network. Figure 1 exhibits the proposed scheme's block diagram.

2.1. Preprocessing

To detect betrayal behaviors, the pre-processing is configured because of time constraints and the shadowing effect [19], [20]. In this step, the image quality is elevated and a set of targeted objects is recognized. Consequently, duplicate frame removal and contrast enhancement are the two steps performed, which are explained as follows,

- Duplicate frame removal: as a part of this subsection, to identify and remove the unchanged frames from the converted frames, the converted frames are compared between them. After removing the duplicates, the frames $\delta^\tau \in \gamma^h$ are expressed as (1),

$$\delta^\tau = \{\delta^1, \delta^2, \delta^3, \dots, \delta^A\} \quad (1)$$

- Contrast enhancement: by utilizing the SE-CLAHE algorithm, the contrast of the remaining frames δ^r is enhanced. The histogram technique that is utilized to augment the structural clarity of the objects in the image is the CLAHE [21]. Only this model among the number of enhancing algorithms has the higher capability to elevate the foggy image's visibility level. However, the incorrect hyperparameter (i.e., clip-limit) selection that diminishes the image quality was utilized by the conventional CLAHE. So, the Shannon entropy values for the histogram were computed by this research methodology and the parameters are selected grounded on the calculated values. After removing the duplicates, the bunch of frames is initially partitioned into contextual regions and the histogram of the contextual regions is computed. To generate the grey level, the mapping function is then applied to each context in the image. For each grey level in the histogram, the number of pixels in the contextual region is divided.

$$P_{avg} = \frac{\rho_u - \rho_v}{\rho_g} \quad (2)$$

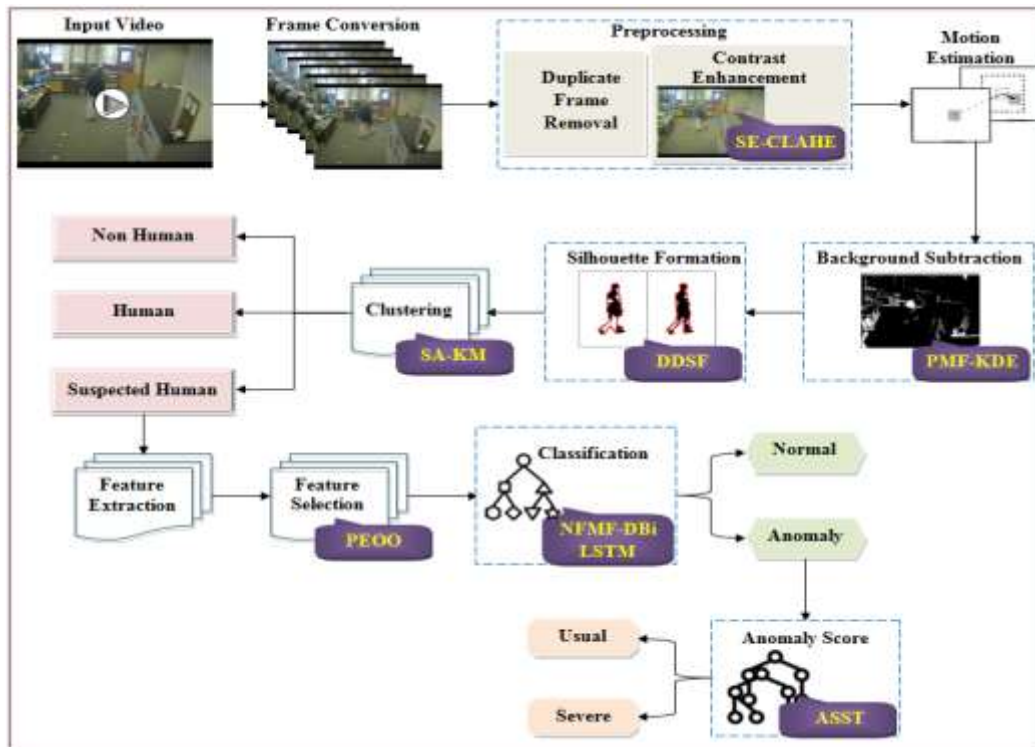


Figure 1. block diagram of the proposed methodology

2.2. Motion estimation

Here, by comparing each frame, the motion of every object is estimated. With the help of using the diamond search algorithm (DSA), the comparison is executed. The images' center-biased motion characteristics are utilized by the diamond search. A large diamond search pattern (LDSP) is utilized recurrently at the image's start [22], [23]. Also, to achieve the minimum block distortion (MBD), the '9' checking points are tested. The MBD point is re-positioned as the centre point to form a new LDSP. After that, the LDSP is changed to the SDSP search pattern. The MBD point detected here is the motion vector's final solution that points to the best matching block. The estimated motion for the object is denoted as $(\delta^v)_{esti}$.

2.3. Background subtraction

Here, by utilizing the PMF-KDE technique, the background for motion-estimated objects $(\delta^v)_{esti}$ is removed. For separating the foreground and background region, the conventional Kernel density estimation helps. It may also assist in the future phase of FE. It can accurately segregate the moving objects by pondering the density betwixt the borderline of the foreground and background region. However, the KDE is centered on choosing the correct bandwidth. If the incorrect bandwidth is chosen then it distorts the data. So, probabilistic matrix factorization based on KDE is the research methodology here. For subtracting

background and foreground, the Gaussian kernel is used here. The bandwidth parameter controls the kernel. Given this kernel formation, the density Φ^v is estimated at an object within a group of the object and it is derived as (3):

$$\Phi^v = \sum_{i=1}^V \kappa(Y - (\delta^v)_{esti}; H) \quad (3)$$

2.4. Silhouette function

For gathering depth information about the object, a silhouette is formed for Φ^v using dirac depth silhouette function (DDSF) [24], [25]. The normal thing with a suspicious appearance was detected by the silhouette's assistance. Depth silhouettes can be seen as a natural extension of binary silhouettes when depth information is presented. To fill the pixels of binary silhouettes with depth, range information is utilized in-depth silhouettes. Therefore, depth silhouettes can register complex body poses devoid of the need to track feature points. However, only the discrete points of the image pixel are considered by the depth silhouette. Continuous points were not considered. Hence, a novel Dirac delta function was proposed in this paper. By assigning a depth value extracted as of a depth map to each foreground pixel, a depth silhouette is engendered as of the binary silhouette. This model can be stretched to handle depth data in the following ways: let $\phi_l \in \Phi^v$ consider the sequence of the depth silhouette of object's each part. It is compacted by generating a depth sampler Γ_l . The depth exemplar's each pixel $\Gamma_l(L_1, L_2)$ is a histogram that signifies the depth values' possible distribution in the silhouette (L_1, L_2) pixels. Consequently, it can be computed as (4):

$$\Gamma_l(L_1, L_2, \mu) = \lambda \sum_{i=1}^{I_{fin}} \Delta_{kro}[\mu_{fun}(\phi_l(L_1, L_2)), \mu] + \int_{-\infty}^{\infty} f(\phi_l, \Gamma_l) \quad (4)$$

2.5. Clustering

To perform non-human, human, and suspected human clustering from Ψ , the sorting and average-based K-means (SA-KM) clustering algorithm is utilized subsequent to the silhouette function [26]. Generally, to group the pixels into the category of human, non-human, and suspected human in the image, K-means is employed. For every new pixel coming in, it computes the distances betwixt pixels. However, the outlier problem still exists in the prevailing models. A sorting and median method is proposed in this paper to sort out this issue. The steps include,

- Initialize all the pixels d^n , which are presented in the silhouette formation as (5):

$$d^n = \{d^1, d^2, d^3, \dots, d^N\} \text{ where, } \alpha = 1, 2, 3, \dots, N \quad (5)$$

here, the number of pixels is signified as α .

- Arrange all the pixels in ascending order.
- To form three cluster groups $C_j \in \Psi$, compute the average of all pixels as:

$$C_j = \frac{d^1 + d^2 + d^3 + \dots + d^N}{3} \text{ where, } j = 1, 2, 3 \quad (6)$$

here, the number of the cluster group is notated as j .

- Choose '3' random points η_j as of '3' separate groups.
- Assign a pixel d_α to the nearest random point.
- Using Euclidean distance, calculate the distance (D_{eucli}) betwixt the data points and random value. It can be articulated as (7):

$$D_{eucli} = \sqrt{\sum_{j=1}^3 \sum_{\alpha=1}^N (\eta_j - d_\alpha)^2} \quad (7)$$

- Choose a cluster for pixels where the distance betwixt the pixel and random value is identical.
- Consider an unsimilar pixel to the subsequent cluster.

2.6. Feature extraction

FE is the process of converting the raw data into numerical features that could be processed when conserving the data in the original data set. The feature-extracted data generated better outcomes than the raw data while performing classification with machine learning [27]. Hence, the angle betwixt joints, floor clip plane, height, width, velocity, and acceleration, the difference between different frames, mean, standard

deviation, skewness, variance, maximal Lyapunov exponent, correlation dimension, motion, size, texture, are the primary features that are extracted. It can be articulated as (8):

$$\aleph_{sus(ex)}^m = \{\aleph_{sus(ex)}^1, \aleph_{sus(ex)}^2, \dots, \aleph_{sus(ex)}^M\} \quad (8)$$

2.7. Feature selection

Utilizing the PEOO algorithm, the vital features are chosen as of $\aleph_{sus(ex)}^m$ after FE. The EO's food behavior in probing for mussels was mimicked by the eurasian oystercatcher optimization (EOO) algorithm [28]. Every bird in the population serves as a search agent. To eat the best mussel (optimal result), the candidate mussel was modified by EO grounded on the best solutions. Regarding size, calories, along with energy, the mussels must be balanced. The account length, energy, and caloric value are considered by the conventional algorithm; all of which are signified by random numbers ranging from 3 to 5. Grounded on the poisson function length, energy, and caloric, this research model chooses those values to prevent this issue. Initially, the extracted features $\aleph_{sus(ex)}^m$ are pondered as the number of EO population. Here, the population's fitness $f_{\aleph_{sus(ex)}^m}$ is derived as (9).

$$f_{\aleph_{sus(ex)}^m} = f\{\aleph_{sus(ex)}^1, \aleph_{sus(ex)}^2, \dots, \aleph_{sus(ex)}^M\} \quad (9)$$

Here, the function of fitness is notated as f . The fitness evaluation is grounded on classification accuracy. Balancing their energy along with calories as of the mussels is the primary goal of EO. A direct relationship betwixt the energy and calories is available in the mussels. As the length increases, the calories and time needed to open mussels increase. Consequently, a huge amount of energy is needed as of EO. During the search process, the EO's behavior is depicted in (10) and (11).

$$E_{fin} = t + E_{current} + \ell * r * (\aleph_{sus(ex)}^{best} - \aleph^{i-1}) \quad (10)$$

$$\aleph^i = \aleph^{i-1} * \xi \quad (11)$$

2.8. Classification

In this section, the selected features \aleph_n are given as input to the NFMMF-DBiLSTM network. Here, human behavior was classified with the help of bidirectional long short-term memory (BiLSTM) [29], [30]. It efficiently analyzed the sentimental behaviors more than the other networks. Since the algorithm should memorize a longer sequence of prior data, more time was required for training. Consequently, in this research methodology, Newton-centered removal of longer memorized values is used. Additionally, the Monotonic activation function is considered to expedite the research. Firstly, the chosen features are given into the input layer that gathers the features. They are further sent to the NFMMF-DBiLSTM layers in a forward and backward manner. The bidirectional layer's output b_l is signified as (12).

$$b_l = \begin{cases} \vec{h_i} t - T_{tot} - 1 \\ \leftarrow h_i t - 1_{tot} - T \end{cases} \quad (12)$$

An output sequence $\vec{h_i}$ that is estimated as of a positive input sequence time from $t - T$ to $t - 1$ was produced by the forward layer NFMMF-DBiLSTM. Whereas, the reversed copy of the forward NFMMF DBiLSTM is the output of the backward $\leftarrow h_i$ along with the input sequence from $t - 1$ to $t - T$. Therefore, by summarizing the forward and backward direction's output, the output layer delivered the output and it is articulated as (13).

$$O_i = \xi(\vec{h_i} + \leftarrow h_i) \quad (13)$$

Forget gate: the data that is relevant to keep as of the prior cell state was decided by the forget gate. It can be articulated as (14),

$$F_g = g(\omega_F[h_{i-1}, \aleph_i] + B_F) \quad (14)$$

where, the hidden layer's weight is denoted as ω_F , the bias vector of NFMF-DBiLSTM cell is signified as B_F , and F_g is the output forget gate.

Input gate: the data relevant to update the NFMF-DBiLSTM unit's current cell state was decided by the input gate. It may be articulated as (15) and (16).

$$I_g = g(\omega_I[h_{i-1}, \mathfrak{I}_i] + B_I) \quad (15)$$

$$g = \frac{1}{1+e^{-i}} \quad (16)$$

Memory cell: by the '2' internal gates, it saved the gathered data and controlled the recurrence of data. By computing the level of memory utilizing newton centered removal technique, the unwanted memory is removed as (17) and (18),

$$\vartheta_i = \frac{\sum_{i=1}^N (h_i^{pre} - \mathfrak{I}_i^{pre})^2}{\sigma_i^2} \quad (17)$$

$$\vartheta_i^{mem} = \begin{cases} \text{if } \vartheta_i > \vartheta_{i_{req}} & \text{then remove the data} \\ \text{if } \vartheta_i < \vartheta_{i_{req}} & \text{then consider the data} \end{cases} \quad (18)$$

2.9. Anomaly score

Here, by utilizing the angle singular spectrum transformation (ASST) approach, the anomaly score is detected for identifying the severity of human behaviors. The technique initially designed for the analysis of time series is named singular spectrum transformation (SST)). The computed anomaly score varies greatly when the major motion patterns are extracted as of the entire human movement at diverse times [31]. It is essential to assess the long-term history of human motion in repetitive tasks over long periods of time to attain the correct anomaly score [32]. The human body's angle of movement between every frame is also very vital; hence this research model is considered. To ponder a more long-term history of human motion, a new computation model of anomaly scores is computed at a time v as (19).

$$S(v) = \frac{\sum_{i=1}^N 1 - \|O_i^{traj(v-(i-1)\Omega)^T} O_i^{test(v)}\|_2}{N} + \frac{\Sigma_{tip} + \Sigma_{elbow}}{2} \quad (19)$$

3. RESULTS AND DISCUSSION

3.1. Database description

128 hours of video footage are encompassed in the UCF-crime dataset. 1,900 uncut real-world surveillance videos of abuse, assault, arrest, road accident, arson, robbery, burglary, shooting, explosion, stealing, vandalism, fighting, and shoplifting are comprised here. As of this, 80% was used for training and 20% was used for testing. Figure 2 shows some sample images of various surveillance videos viz., Figures 2(a)-(c) exhibit some of the frame converted images, their corresponding contrast-enhanced images, and the background-subtracted images respectively.

3.2. Performance analysis of enhancement

Regarding peak signal to noise ratio (PSNR), mean square error (MSE), together with structural similarity index (SSIM), the proposed SE-CLAHE's performance analysis is validated. The outcomes obtained as of these metrics are analogized with the prevailing models like CLAHE, double plateau histogram equalization (DPHE), adaptive histogram equalization (AHE), and histogram equalization (HE).

The proposed model and the prevailing model's performance comparison are depicted in Figure 3. Since the PSNR value increases, the proposed model's quality improves. Figure 3(a) exhibits that the PSNR of 35 db was attained by the proposed SE-CLAHE; whereas the PSNR of prevailing methodologies is 29 db for CLAHE, 25 db for DPHE, and 24 db for AHE. A higher PSNR value was attained by the proposed scheme when those values are contrasted with the existing models. Likewise, the proposed SE-CLAHE's performance was improved by a lower MSE value. Figure 3(b) exhibits that the proposed model's MSE is 16.98, which is lower when analogized with the prevailing models like CLAHE, DPHE, AHE, and HE. Similarly, Figure 3(c) illustrates that the proposed model's SSIM is 0.9789 and the prevailing models attain an SSIM in the range of 0.901 to 0.9506. The proposed model has a higher SSIM value when analogized with the prevailing schemes. It is finalized from all the metrics that better performance was unveiled by the proposed SE-CLAHE in preprocessing.

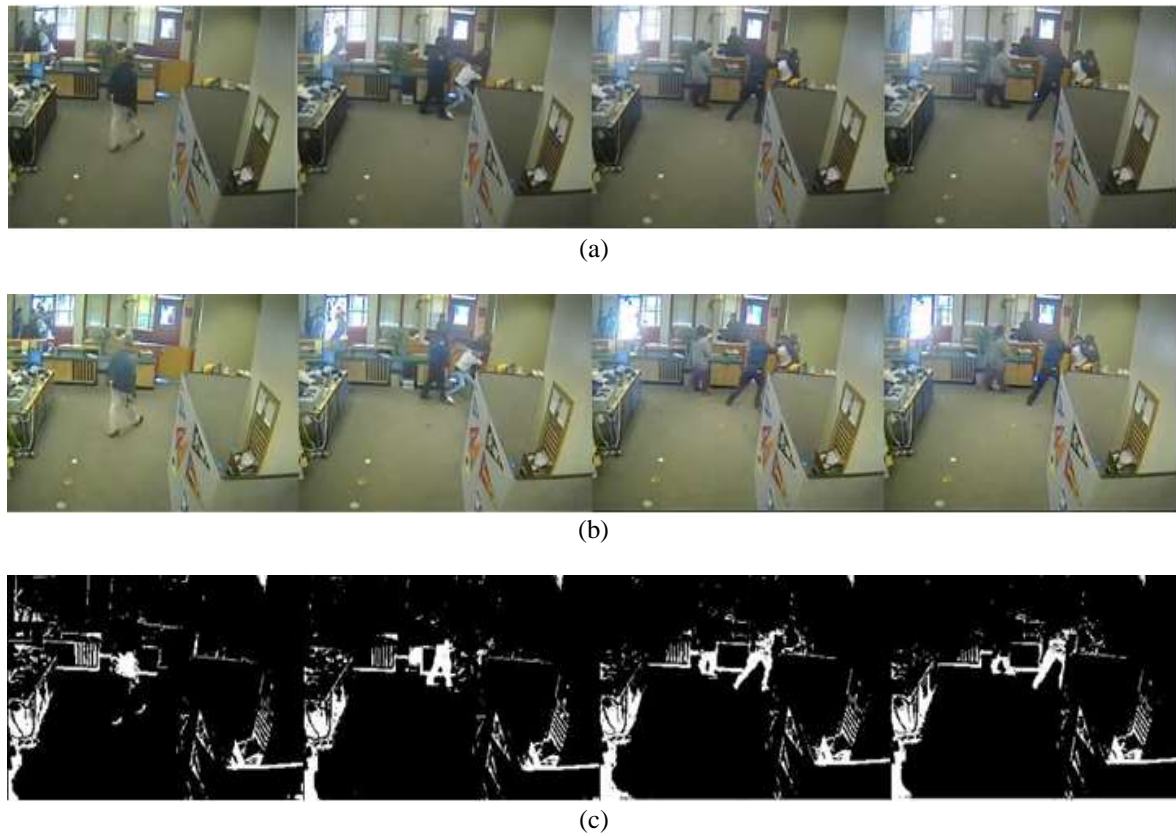


Figure 2. Sample images of various surveillance videos; (a) frame-converted images, (b) contrast-enhanced images, and (c) background subtraction

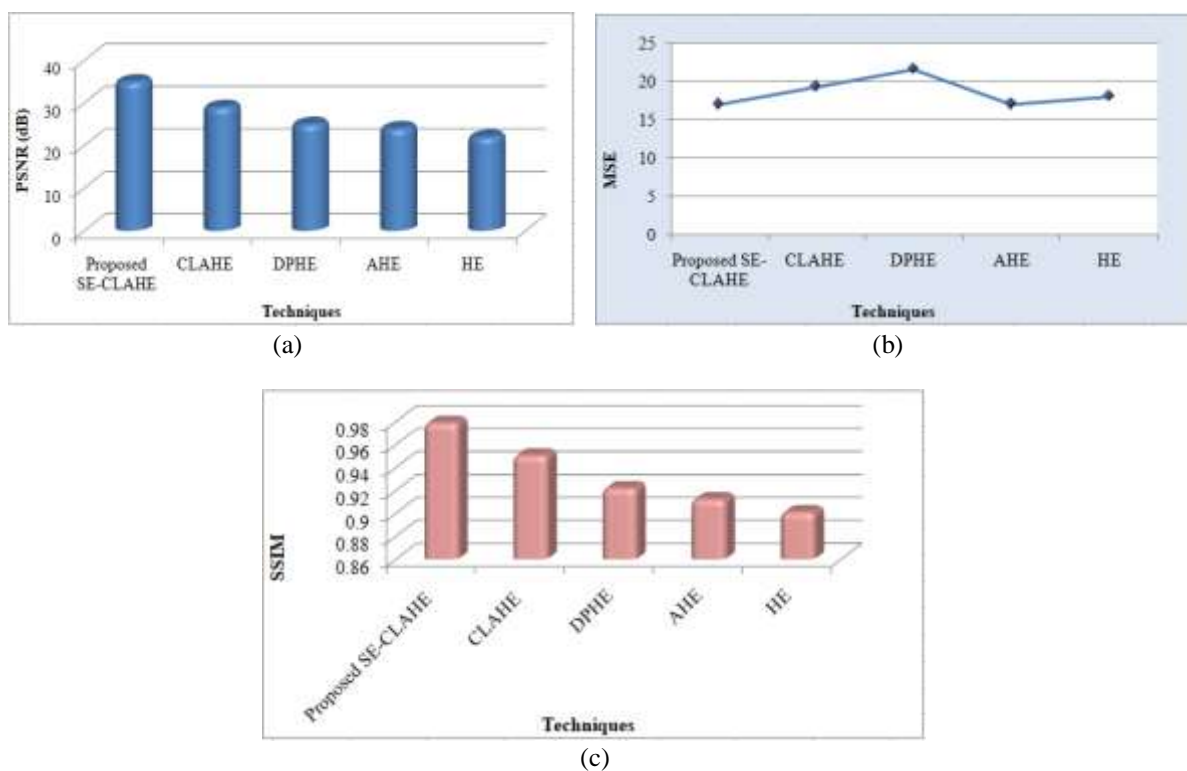


Figure 3. Performance analysis of proposed model and existing model: (a) PSNR, (b) MSE, and (c) SSIM

3.3. Performance analysis of feature selection

The proposed PEOO's performance is compared with the prevailing methodologies like EOO, aquila optimization (AO) algorithm, gannet optimization (GO), and salp swarm optimization (SSO) algorithm to prove the developed model's superiority. The convergence analysis's outcomes are illustrated in Figure 4. In this case, the fitness function is classification accuracy. As of the graph, the proposed PEOO's efficiency attained for 10 iterations, 20 iterations, and 30 iterations are 90%, 93.21%, and 94.7%. Nevertheless, the prevailing EO's fitness for 10 iterations and 20 iterations are 87% and 89.76% respectively. Also, the remaining models are analogized with the proposed model. These values are found to be lesser than the prevailing models. The analysis exhibited that when analogized with prevailing models, the proposed model performs better.

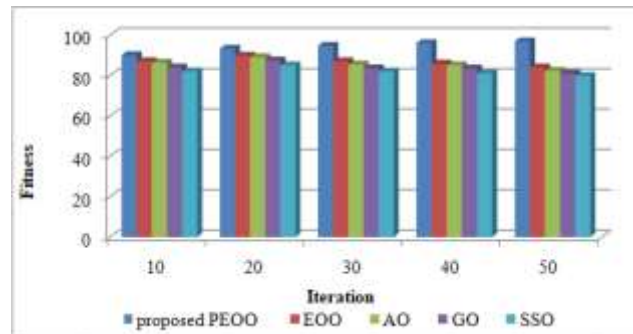


Figure 4. Convergence analysis

4. CONCLUSION

Human AD was proposed in this work. Frame conversion, preprocessing, motion estimation, background subtraction, silhouette function, clustering, FE, feature section, and classification are performed in this proposed model. The experimental analysis is executed after performing all the steps. The proposed NFMF-DBiLSTM's performance is analyzed as well as analogized with the prevailing techniques. The final outcomes exposed that an accuracy of 96.98% and a computational time of 6,023 ms were attained by the proposed model. Likewise, for all the metrics like precision, recall, F-measure, false rejection rate (FRR), false negative rate (FNR), PSNR, MSE, and training time, the proposed model achieved better results. So, it is concluded as of the results of all metrics that the proposed model is highly efficient than the prevailing models. The work is not interiorly concentrated on the object when they are in a group. Thus, for considering the people when they are in the group, future work will be extended with some advanced techniques.





REFERENCES

- [1] Z. K. Abbas and A. A. Al-Ani, "Anomaly detection in surveillance videos based on H265 and deep learning," *International Journal of Advanced Technology and Engineering Exploration*, vol. 9, no. 92, pp. 910–922, Jul. 2022, doi: 10.19101/IJATEE.2021.875907.
- [2] F. Abdullah, M. Javeed, and A. Jalal, "Crowd anomaly detection in public surveillance via spatio-temporal descriptors and zero-shot classifier," in *2021 International Conference on Innovative Computing (ICIC)*, Nov. 2021, pp. 1–8, doi: 10.1109/ICIC53490.2021.9693003.
- [3] C. . Amrutha, C. Jyotsna, and J. Amudha, "Deep learning approach for suspicious activity detection from surveillance video," in *2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*, Mar. 2020, pp. 335–339, doi: 10.1109/ICIMIA48430.2020.9074920.
- [4] U. Bedekar and G. Bhatia, "A novel approach to recommend skincare products using text analysis of product reviews," in *Lecture Notes in Networks and Systems*, vol. 191, 2022, pp. 241–251.
- [5] D. Chen, P. Wang, L. Yue, Y. Zhang, and T. Jia, "Anomaly detection in surveillance video based on bidirectional prediction," *Image and Vision Computing*, vol. 98, p. 103915, Jun. 2020, doi: 10.1016/j.imavis.2020.103915.
- [6] K. Doshi and Y. Yilmaz, "Continual learning for anomaly detection in surveillance videos," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jun. 2020, pp. 1025–1034, doi: 10.1109/CVPRW50498.2020.00135.
- [7] R. J. Franklin, Mohana, and V. Dabbagol, "Anomaly detection in videos for video surveillance applications using neural networks," in *2020 Fourth International Conference on Inventive Systems and Control (ICISC)*, Jan. 2020, pp. 632–637, doi: 10.1109/ICISC47916.2020.9171212.
- [8] M. Gnouma, R. Ejbali, and M. Zaied, "Video anomaly detection and localization in crowded scenes," in *Advances in Intelligent Systems and Computing*, vol. 951, 2020, pp. 87–96.
- [9] M. Gong, H. Zeng, Y. Xie, H. Li, and Z. Tang, "Local distinguishability aggrandizing network for human anomaly detection," *Neural Networks*, vol. 122, pp. 364–373, Feb. 2020, doi: 10.1016/j.neunet.2019.11.002.





- [10] D. Koshti, S. Kamoji, N. Kalnad, S. Sreekumar, and S. Bhujbal, "Video anomaly detection using inflated 3D convolution network," in *2020 International Conference on Inventive Computation Technologies (ICICT)*, Feb. 2020, pp. 729–733, doi: 10.1109/ICICT48043.2020.9112552.
- [11] M. K. Tripathi and Shivendra, "Neutrosophic approach based intelligent system for automatic mango detection," *Multimedia Tools and Applications*, vol. 83, no. 14, pp. 41761–41783, 2024, doi: 10.1007/s11042-023-17037-7.
- [12] T. Saba, "Real time anomalies detection in crowd using convolutional long short-term memory network," *Journal of Information Science*, vol. 49, no. 5, pp. 1145–1152, Oct. 2023, doi: 10.1177/01655515211022665.
- [13] K. K. Santhosh, D. P. Dogra, and P. P. Roy, "Anomaly detection in road traffic using visual surveillance," *ACM Computing Surveys*, vol. 53, no. 6, pp. 1–26, Nov. 2021, doi: 10.1145/3417989.
- [14] M. K. Tripathi, M. Neelakantappa, A. N. Kaulage, K. V. Nabilal, S. N. Patil, and K. D. Bamane, "Breast cancer image analysis and classification framework by applying machine learning techniques," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 11, no. 3, pp. 930–941, 2023.
- [15] A. Sharma and N. Varshney, "Identification and detection of abnormal human activities using deep learning techniques," *European Journal of Molecular and Clinical Medicine*, vol. 7, no. 4, pp. 408–417, 2020.
- [16] H. Shin, K. Na, J. Chang, and T. Uhm, "Multimodal layer surveillance map based on anomaly detection using multi-agents for smart city security," *ETRI Journal*, vol. 44, no. 2, pp. 183–193, Apr. 2022, doi: 10.4218/etrij.2021-0395.
- [17] W. Ullah, A. Ullah, I. U. Haq, K. Muhammad, M. Sajjad, and S. W. Baik, "CNN features with bi-directional LSTM for real-time anomaly detection in surveillance networks," *Multimedia Tools and Applications*, vol. 80, no. 11, pp. 16979–16995, May 2021, doi: 10.1007/s11042-020-09406-3.
- [18] W. Ullah, A. Ullah, T. Hussain, Z. A. Khan, and S. W. Baik, "An efficient anomaly recognition framework using an attention residual LSTM in surveillance videos," *Sensors*, vol. 21, no. 8, p. 2811, Apr. 2021, doi: 10.3390/s21082811.
- [19] S. W. Yahaya, A. Lotfi, and M. Mahmud, "Towards a data-driven adaptive anomaly detection system for human activity," *Pattern Recognition Letters*, vol. 145, pp. 200–207, May 2021, doi: 10.1016/j.patrec.2021.02.006.
- [20] C. V. K. N. S. N. Moorthy, M. K. Tripathi, S. Joshi, A. Shinde, T. K. Zope, and V. U. Avachat, "SEM and TEM images' dehazing using multiscale progressive feature fusion techniques," *Indonesian Journal of Electrical Engineering and Computer Science (IJECS)*, vol. 33, no. 3, pp. 2007–2014, 2024, doi: 10.11591/ijeecs.v33.i3.pp2007-2014.
- [21] F. Yang, Z. Yu, L. Chen, J. Gu, Q. Li, and B. Guo, "Human-machine cooperative video anomaly detection," *Proceedings of the ACM on Human-Computer Interaction*, vol. 4, no. CSCW3, pp. 1–18, Jan. 2021, doi: 10.1145/3434183.
- [22] M. K. Tripathi and Shivendra, "Improved deep belief network for estimating mango quality indices and grading: a computer vision-based neutrosophic approach," *Network: Computation in Neural Systems*, pp. 1–29, 2024, doi: 10.1080/0954898X.2023.2299851.
- [23] J. Ramkumar, C. Karthikeyan, E. Vamsidhar, and K. N. Dattatraya, "Automated pill dispenser application based on IoT for patient medication," *EAI/Springer Innovations in Communication and Computing*, pp. 231–253, 2020, doi: 10.1007/978-3-030-42934-8_13.
- [24] K. N. Dattatraya, K. R. Rao, and D. Satish Kumar, "Architectural analysis for lifetime maximization and energy efficiency in hybridized WSN model," *International Journal of Engineering and Technology(UAE)*, vol. 7, pp. 494–501, 2018, doi: 10.14419/ijet.v7i2.7.10870.
- [25] K. N. Dattatraya and S. Ananthakumaran, "Energy and trust efficient cluster head selection in wireless sensor networks under meta-heuristic model," *Lecture Notes in Networks and Systems*, vol. 444, pp. 715–735, 2022, doi: 10.1007/978-981-19-2500-9_53.
- [26] K. N. Dattatraya, S. Ananthakumaran, and K. V.D. Kiran, "Optimal cluster head selection in wireless sensor network via improved moth search algorithm," *Artificial Intelligence in Information and Communication Technologies, Healthcare and Education: A Roadmap Ahead*, pp. 95–108, 2022, doi: 10.1201/9781003342755-11.
- [27] M. K. Tripathi, P. K. Reddy, M. Neelakantappa, C. V. Andhare, and Shivendra, "Identification of mango variety using near infrared spectroscopy," *Indonesian Journal of Electrical Engineering and Computer Science (IJECS)*, vol. 31, no. 3, pp. 1776–1783, 2023, doi: 10.11591/ijeecs.v31.i3.pp1776-1783.
- [28] M. Z. Zaheer, A. Mahmood, M. H. Khan, M. Astrid, and S.-I. Lee, "An anomaly detection system via moving surveillance robots with human collaboration," in *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, Oct. 2021, pp. 2595–2601, doi: 10.1109/ICCVW54120.2021.00293.
- [29] C. H. V. K. N. S. N. Moorthy, M. K. Tripathi, N. P. Bhosle, V. A. Suryawanshi, P. A. Kadam, and S. A. Bahir, "Investigating advanced and innovative non-destructive techniques to grade the quality of *Mangifera Indica*," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 1, pp. 299–309, 2024.
- [30] X. Zhang, A. Ghimire, S. Javed, J. Dias, and N. Werghi, "Robot-person tracking in uniform appearance scenarios: a new dataset and challenges," *IEEE Transactions on Human-Machine Systems*, vol. 53, no. 3, pp. 549–559, 2023, doi: 10.1109/THMS.2023.3247000.
- [31] M. K. Tripathi, S. Hasini, M. Homamalini, and M. Neelakantappa, "Pothole detection based on machine learning and deep learning models," in *Proceedings of 3rd International Conference on Advanced Computing Technologies and Applications, ICACTA 2023*, 2023, pp. 1–9, doi: 10.1109/ICACTA58201.2023.10393510.
- [32] T. Xiang *et al.*, "Exploiting structural consistency of chest anatomy for unsupervised anomaly detection in radiography images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024, doi: 10.1109/TPAMI.2024.3382009.

BIOGRAPHIES OF AUTHORS







Dr. Sanjeevkumar Angadi     received Ph.D. from Visvesvaraya Technological University, Karnataka, India in 2023 and M. Tech in Computer Science and Engineering from Visvesvaraya Technological University, Karnataka, in 2012. At present he is working as Associate Professor Head of Department in the Department of Computer Science and Engineering at PCET's NMVPM Nutan College of Engineering and Research, Talegaon Dabhade, Pune, India. He has published 25+ research papers in international journals and conferences and granted two patents. His research interest includes computer vision, image processing, machine learning, data mining, and natural language processing. He can be contacted at email: angadi.sanjeevkumar@gmail.com.







Dr. Chellapilla V. K. N. S. N. Moorthy     is working as Director RD, Vasavi College of Engineering, Hyderabad, Telangana, India. He received Master of Technology both in the fields of Computer Science Engineering and Heat Power Refrigeration and Air Conditioning. He received doctoral degree for research from GITAM University, and pursuing his doctoral degree in the field of machine learning too. He has research experience with a total research grant of 424.46 K USD from DST, government of India, more than 50 research publications, international research collaborations, awards and patents to his credit. His thrust areas of research include cognitive science, AI and ML, and computational fluid dynamics. He can be contacted at email: krishna.turbo@gmail.com.







Dr. Mukesh Kumar Tripathi     received a Ph.D. in Computer Science and Engineering from VTU, Belagavi. He has 15 years of teaching and administrative experience. He has supervised and co-supervised more than five masters and 20 B.E. students. He is working as an assistant professor with the Department of Computer Science and Engineering, Vardhaman College of Engineering, Hyderabad, India. He is authored or co-authored more than twenty publications and more than 300 citations. His research interests include soft computing, machine learning, intelligent systems, image processing, and Hyperspectral. He can be contacted at email: mukeshtripathi016@gmail.com.







Dr. Bhagyashree Ashok Tingare     has completed her Ph.D. in Computer Engineering with Augmented Reality and Virtual Reality Specialization from Shri Jagdishprasad Jhabarmal Tibrewala University, Rajasthan. She also received a B.E. and M.E. in Computer Engineering from D. Y. Patil College of Engineering, Akurdi, Pune, Maharashtra India. She has 9 years of teaching and Industry experience. She has supervised and co-supervised more than 12 B.E. Projects. She has authored or co-authored more than 30 publications. She is working as an Assistant professor with the Department of Artificial Intelligence and Data Science, D. Y. Patil College of Engineering, Akurdi, Pune, Maharashtra India. Her research interests include natural language processing and artificial intelligence. She can be contacted at email: bhagyashreetingare@gmail.com.



Dr. Sandeep Uddhavrao Kadam     working as Vice President, WeshineTech pvt ltd. He is having 20+ years in academics and industry. His research area is cyber security and software engineering, include soft computing, machine learning, intelligent systems, image processing, and hyperspectral. He is Ph.D. research Co-Supervisor in Computer Engineering. He is authored or co-authored more than twenty-five publications and more than 100 citations. He can be contacted at email: sandipkadam@gmail.com.



Dr. Kapil Misal     received master of Computer Application from Dr. Babasaheb Ambedkar Marathwada University. He has accomplished his Ph.D. degree from Dr. Babasaheb Ambedkar Marathwada University. He has published more than 6 papers in international and national journals and conferences. He has proposed research project funded by Biotechnology Industry Research Assistance Council. Currently, he is working as an Associate Professor in Department of Master of Computer Application at Trinity Academy of Engineering Pune. He has 10 years of teaching experience and 1 years of industry experience. His research interest includes machine learning and computer vision. He can be contacted at email: misal.kapil@gmail.com.