INTELLIGENT DEEP LEARNING ENABLED CROWD DETECTION AND CLASSIFICATION MODEL IN REAL TIME SURVEILLANCE VIDEOS

S. Sivachandiran

[1]Research Scholar, Department of Computer Science & Engineering, Annamalai University,
Tamil Nadu, India.
sivachandiran.s@gmail.com

Dr. K. Jagan Mohan

[2]Associate Professor, Department of Information Technology, Annamalai University,
Tamil Nadu, India.
aucsejagan@gmail.com

Dr. G. Mohammed Nazer

[3]Professor, Department of Computer Science, RAAK College of Arts and Science, Puducherry.
kgmohammednazer@gmail.com

Abstract

Recently, security surveillance applications exploited the computer vision based detection and tracking approaches to improve the safety and comfort of humans. A major concern in real time surveillance video tracking is the process of identifying the human crowd behavior and classifying them. It finds useful to alert the crow in case of any disasters and unpredicted events. The investigation of human behavior in crowded surveillance videos is an essential and crucial area of research. The recent advances in Artificial Intelligence (AI) and deep learning (DL) models can be employed for determining the crowd behavior analysis in surveillance videos. With this motivation, this article focuses on the design of intelligent deep learning enabled crowd behavior detection and classification (IDL-CBDC) model in real time surveillance videos. The goal of the IDL-CBDC technique is to detect the crowd and classify it into four classes namely marriage, political, school, and college. Primarily, the IDL-CBDC technique performs preprocessing in two levels namely adaptive median filtering (AMF) technique and contrast enhancement (CE) approach. Besides, a deep instance segmentation approach using PSPNet-101 model is used for the segmentation of input video frames into crowds. Moreover, the black widow optimization (BWO) with residual network (ResNet50) model is applied for the crowd detection and classification process. The design of BWO algorithm helps to properly adjust the hyperparameter values such as learning rate, batch size, number of epochs, and number of hidden layers. In order to ensure the improved performance of the IDL-CBDC technique, a set of simulations take place using an own dataset, gathered from public places. Extensive comparative result analysis reported the supremacy of the IDL-CBDC technique over the other techniques.

***Keywords*: Video surveillance, Real time videos, Deep learning, Object detection, Crowd detection, Parameter tuning.**

1. Introduction

In recent times, security surveillance system has employed visual-based tracking and detection technologies for enhancing safety and convenience for human beings [1]. Human tracking and detection systems are important topics in a surveillance scheme. Moving object extraction and Human recognition are the two major parts of human detection method. Human recognition detects an object as human or nonhuman, and object is extracted from the background through moving object extraction that defines the relevant position and size of the objects in an image [2]. The tracking method is capable of predicting the position after and during occlusion since the tracked human or object is occluded probably by other objects while tracked. Typically, Surveillance systems used two types of cameras: active cameras and fixed cameras [3]. An active camera takes appropriate limited field of view (FOV) since it could perform pan–tilt for retaining the targeted objects within the camera scene, whereas the fixed cameras have the benefits of being lower cost but come with FOV. Additionally, the latter has a good solution as it could implement zoom in or out operation [4]. In general, a tracking scheme on active cameras considers the temporal variation to extract moving objects. In this way, it is essential to wait for the camera to be strong enough to operate the image [5]. In another word, the moving camera captures blurred image and extract the moving object and background pixels. Then, the active camera is processed discontinuously and non-smoothly. Therefore, a particle filter tracking method is employed for resolving these problems [6]. The codebook algorithm is initially applied for spotting the human as the targeted system, and later the particle filters track the human by the computation of Bhattacharyya distance amongst the color histogram of targeted models with the color histogram frames of the sample particle position. There are several benefits of utilizing color histograms namely scale invariant, effective computation, rotation, tracking of non-rigid objects, and strength to partial occlusion [7].

The major challenge in monitoring and surveillance schemes is to detect human crowd behaviours and supervise the crowd for preventing disaster and unexpected events [8]. Analyzing human behaviour in crowded scenes is the most challenging and important field in recent studies [9]. Conventional visual surveillance system purely depends on manpower to examine videos is ineffective as a result of the massive amount of screens and cameras that need monitoring, human fatigue because of time consumed on long monitoring period, training in what to look for, and paucity of important fore-knowledge and also due to the enormous number of video information i.e., created for every day. This issue necessitates an automatic visual surveillance scheme that could reliably isolate, analyze, detect, alert and identify responders to anomalous events in realtime [10]. An automatic surveillance system seeks to automatically identify human behaviour in crowded scenes, and it has several applications, like traffic monitoring, security, military applications, inspection tasks, sports analysis, robotic vision, pedestrian traffic monitoring, video surveillance, care of the elderly, and infirm.

Pawar and Attar [11] focused on analyzing and studying DL methods for video-based anomalous activity recognition. In this work, the graphical taxonomy has been presented based on level of anomaly detection, kind of anomalies, and anomaly measurements for detecting anomalous activities. Special emphasis has been given on different anomaly detection architectures having DL technique as baseline method. The DL methods from the perspective of realtime processing-and accuracy-oriented anomaly detections are compared to one another. Rezaei and Yazdi [12] examined a deep framework for crowd event detection to recognize 7 behaviour classes in PETS2009 event detection data set. Particularly employs a Conv-LSTM-AE and handcrafted methodology with optical flow images as input for extracting a higher-level depiction of data and performing classification. Afterward accomplishing a latent depiction of input optical flow image sequence in the bottleneck of AE, the framework is split into classifier and AE decoder. Varghese and Thampi [13] presented a method to predict crowd behaviours using a DL architecture and multi-class SVM. They extracted spatio-temporal descriptors using three dimensional CNN based crowd emotion. Gao et al. [14] developed a crowd analysis methodology combining compressive sensing and DL network for detecting violent behaviours. For achieving this goal, a hybrid random matrix (HRM) is created and is demonstrated to fulfill the restricted isometry property. The higher-dimension feature is proposed to a lower-dimension space by using the HRM. Song and Sheng [15] introduced a single-image crowd counting and abnormal behaviours detection using multi-scale GAN networks. The presented model initially developed an embedded GAN method using a regional discriminator and a multi-branch generator to initially create crowd-density map; later the presented method is added to reinforce the generalization capability of the method.

This article presents an intelligent deep learning enabled crowd behavior detection and classification (IDL-CBDC) model in real time surveillance videos. Initially, the IDL-CBDC technique performs preprocessing in two levels namely adaptive median filtering (AMF) technique and contrast enhancement (CE) approach. In addition, a deep instance segmentation approach using PSPNet-101 model is used for the segmentation of input video frames into crowds. Followed by, the black widow optimization (BWO) with residual network (ResNet50) model is applied for the crowd detection and classification process. In order to ensure the improved performance of the IDL-CBDC technique, a set of simulations take place using an own dataset, gathered from public places.

2. The Proposed Model

In this study, a new IDL-CBDC technique has been derived for crowd behavior analysis in real time surveillance videos, which mainly classifies the video frames into four classes namely marriage, political, school, and college. The proposed IDL-CBDC technique comprises preprocessing, AMF based noise removal, contrast enhancement, DIS based segmentation, ResNet50 based feature extraction, softmax layer based classification, and BWO based hyperparameter optimization. Fig. 1 illustrates the overall working process of proposed IDL-CBDC technique.

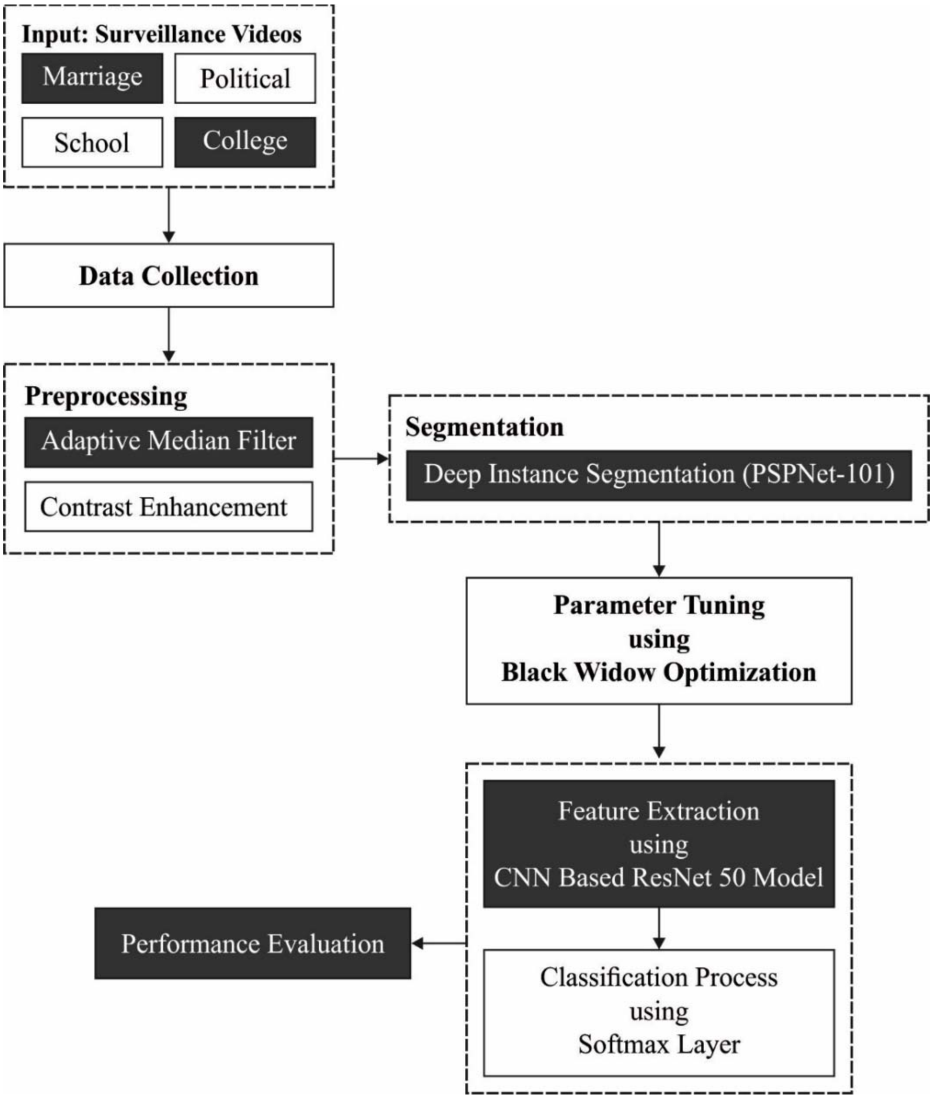
Input: Surveillance Videos
Marriage
Political
School
College
Data Collection
Preprocessing
Adaptive Median Filter
Contrast Enhancement
Segmentation
Deep Instance Segmentation (PSPNet-101)
Parameter Tuning using Black Widow Optimization
Feature Extraction using CNN Based ResNet 50 Model
Classification Process using Softmax Layer
Performance Evaluation

Fig. 1. Overall process of IDL-CBDC technique

2.1. *Pre-Processing*

At the initial stage, the real time surveillance videos are initially separated into a set of frames. Then, the preprocessing is carried out in two stages namely AMF based noise removal and CLAHE based contrast enhancement.

2.1.1. Adaptive Median Filter

A simple MF [16] uses the median of the window for replacing the central pixel assumed as a window. When the center pixel is (Pepper) or (salt), it would be replaced with the middle value of the window. The main disadvantage of standard median filter is that even though the pixel is considered incorrect (except 0 or 255), it is replaced with the median of the window. This would damage the entire visual quality of an image. Additionally, a simple MF could not retain edges. The window is arranged in ascending. Median is the middle value afterward the sort. Thus, the undamaged pixel is replaced with the median value of the window.
Sample window Output

$$\begin{bmatrix} 46 & 64 & 82 \\ 255 & (45) & 52 \\ 64 & 64 & 82 \end{bmatrix} \rightarrow \begin{bmatrix} 46 & 64 & 82 \\ 255 & (82) & 52 \\ 64 & 64 & 82 \end{bmatrix}$$

As mentioned below, when the pixel is destroyed, the impulse noise would be eliminated similarly [17].

Sample window Output

$$\begin{bmatrix} 46 & 64 & 82 \\ 255 & (255) & 52 \\ 64 & 64 & 82 \end{bmatrix} \rightarrow \begin{bmatrix} 46 & 64 & 82 \\ 255 & (82) & 52 \\ 64 & 64 & 82 \end{bmatrix}$$

2.1.2. Contrast Enhancement

The next phase is employing CLAHE on the noise removed image. It has better tractability in selecting local histogram mapping functions. The CLAHE technique splits the image into suitable regions and employs histogram equalization to them. This technique alters the intensity value of an image by using non-linear method for maximizing the contrast for each pixel of an image. The clipping level selection of the histogram decreases the unwanted noise amplification. The clipped pixel is reallocated to all the gray levels. The novel histogram is distinct from the standard histogram since intensity of all the pixels is constrained by user-selectable maximum. Therefore, the CLAHE technique could limit the noise improvement

2.2. *Deep Instance Segmentation*

Indeed, PSPNet is the more commonly known network structure for semantic segmentation [18]. The PSPNet is originally presented for scene parsing. In order to aggregate multiscale context data, a single PPM was presented in a PSPNet. Initially, max pooling is employed for generating feature maps with three distinct pyramid scales as Eq. (1), where, FDS and λ indicate the input, down sampling process by max −pooling, and the strides of the max −pooling layer, correspondingly. The stride of max −pooling layer is attained by Eq. (2), where w and 0 denote the size of the input and output feature map.

$$F_j = DS(F, \lambda_j.)j = 1,2,3 \tag{1}$$

$$\frac{w - \lambda_j}{\lambda_j} + 1 = 0_j \Rightarrow \lambda_j = \frac{w}{o_j} \tag{2}$$

Afterward employing convolutional process to this multiscale feature map, bi-linear interpolation is executed for resizing the feature map as Eq. (3), while W_j^T and b_j indicates the weight and bias of the jth 1×1 convolution layer, correspondingly, $BI(.)$ represents the bilinear interpolation.

$$O_j = BI\left(W_j^T \otimes F_j + b_j\right) j = 1,2,3 \tag{3}$$

Furthermore, the feature map with distinct pyramid scales and the new input are concatenated, and a 1×1 convolutional process is employed for reducing the channel number of the output as Eq. (4), while W_j^T and b_j shows the weight and bias of 1×1 convolutional layer.

$$C = W_{rd}^T\left(concat(F, O_1, O_2, O_3)\right) + b_{rd} \tag{4}$$

Unlike the original PPM, feature maps with three pyramid scales, involving 1, 2, and 6, are created using max pooling [19], in which feature maps with four pyramid scales, involving 1, 2, 3, and 6, are created using the novel PPM.

Additionally, a 1×1 convolutional layer was connected to the concatenation layer for reduction dimension.

Stimulated by the architecture of UNet, a multi-level PSPNet is presented as the decoder. The 1st, 2nd, and 3rd attention gates are employed for generating the attention map of the 5th identity blocks, the 3rd identity blocks, and the 1st convolutional layer, correspondingly. Moreover, to integrate multi-level features, the output of the PPM and the attention gate is densely concatenated as follows.

$$Y_j = concat\left(US\left(C_j, 3\right)M\_output_j\right)j = 1,2,3 \tag{5}$$

2.3. *Crowd Detection and Classification*

During the classification process, the ResNet50 model can be utilized to detect and classify the crowds in the surveillance videos. To characterize the image, we adapt the CNN, $ResNet50$, i.e., deeper networks have fifty layers. The depth of the network is critical for neural networks, however, deep network is harder to train. The architecture of $ResNet50$ facilitate the training of network and permits them to be very deep that resulting in improved efficiency in various tasks. ResNet50 is very deep when compared to simple” counterparts, however, the amount of parameters of this network is very small. A DCNN has led to a sequence of breakthroughs for the classification of images [20]. Several more non-trivial visual detection tasks have benefited considerably from the deep model. When the network depth increases, the accuracy of network rapidly increased and degrades quickly (saturated). Since a deep network has a larger representation power. It is probable for ResNet50 to attain deep models trivially that isn’t worse than the less deep network.

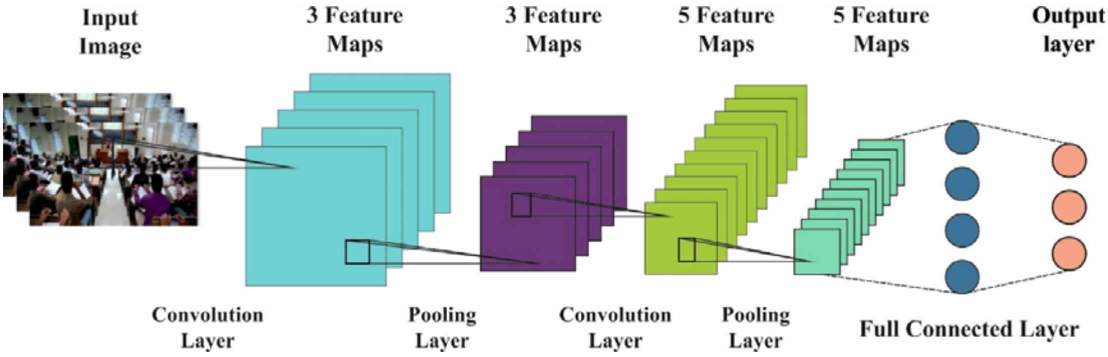

Fig. 2. Layers in CNN Architecture

It can be performed by adding many identity layers, i.e., levels that only skips the signal without any modification. $ResNet50$'s deep level has to forecast the variation among the objective function and output of the preceding layers. They can drive the weight to zero and only skips the signal. Therefore, deep residual learning is an effective technique which makes the networks learn to forecast deviation from previous layer. Fig. 2 showcases the layers in CNN structure.

The model takes an image and produces a caption, encoded as a series of $1-K$ codewords.

$$y=\{y_1,y_2,\cdots,y_c\}, y_j\in R^K, \tag{6}$$

Whereas K represents the size of the dictionary and c denotes the caption length. They utilize CNN, $ResNet50$, to attain set annotation vectors such as feature vectors. The extractor produces L-vector, each containing a D-dimension representations of an image.

For properly adjusting the hyperparameter values such as learning rate, batch size, number of epochs, and number of hidden layers, the BWO algorithm can be employed. In general, the BW is the spider which comes under the subclass Latrodectus and they aren't very popular due to venomous characteristics. Generally, the phases of BW spider are categorized into 5 major classes as Procreation rate, generation of the population initialization, mutation rate, convergence rate, and Cannibalism rate i.e., given below. Now, in the BWO method, the widow is estimated as a possible solution for all the optimization problems [21]. For assumption, the dimension optimization issue is represented as M^{VAR}; next, the array of all the BWs are denoted as $1\times M^{VAR}$.

$$WB_W=[Q_1,Q_2,\ldots Q_{M_{VAR}}] \tag{7}$$

From Eq. (7), the parameter $[Q_1,Q_2,\ldots Q_{M_{VAR}}]$ indicates the overall amount of floating points and WB_W signifies the BW. Then, the fitness values are estimated by creating the objective function.

$$Obj: F(WB_W)=F[Q_1,Q_2,\ldots Q_{M_{VAR}}]. \tag{8}$$

Next, the matrix with the size $M^{VAR}\times M^{POP}$ is utilized for generating the candidate widow matrix. The procedure of matting is performed for randomly selecting the parent widows. Considering the fact that BW doesn't depend on one another and they begin mating with one another for developing a new generation. Now, in this procedure, only the mating BW is on the web wherein they isolate itself from others. At the same time, over a thousand eggs are laid by the female spider. Amongst this egg, only some egg comprising spider is hatched stronger when the other baby spiders are weaker. Hence, an alpha array is established to reproduce an off-spring that comprises of random number is expressed by:

$$A_1=(1-\delta)Q_2+\delta Q_1, \tag{9}$$

$$A_2=(1-\delta)Q_1+\delta Q_2. \tag{10}$$

Thus, from Eqs. (9) and (10), the offspring and the spider is denoted by A_1 and A_2, Q_1 and Q_2 correspondingly.

In this stage, the rate of cannibalism is of 3 distinct stages such as sexual cannibalism, mother BW, and sibling cannibalism ate the young BW [22]. In sibling cannibalism, the strong young BW spiders kill their own sibling i.e., weaker. The 3rd kind of cannibalism includes eating the younger spider by the mother BW. Therefore, the fitness value defines the power of the spider whether it is stronger or weaker.

In this stage, the mutepop window is randomly chosen from the whole population. Besides this, the mutation rate = is attained by estimating the mutepop. In the BWO method, three major features are to be deliberated when estimating the convergence rate. The factor includes the iteration number is predefined, the optimum value is constant for numerous iterations. The process included in the BWO Algorithm is given below. The BWO method derives a fitness function to obtain better classification accuracy. It defines a positive integer to characterize the good efficiency of the candidate solution. In our work, the minimalization of the classification error rate is regarded as the FF. The optimum solution has a minimal error rate and the worst solution obtains an improved error rate.

$$fitness(x_i) = ClassifierErrorRate(x_i) = \frac{numerb\ of\ misclassified\ instances}{Total\ number\ of\ instances} * 100 \qquad (11)$$

Pseudocode: Black Widow Optimization (BWO) Algorithm

```
Input: procreation rate, mutation rate, cannibalism rate, maximal iteration Output: optimum solution
start
Population Initialization;
reproduction number $\boldsymbol{r_n}$ estimation-based procreation rate;
optimum selection of $\boldsymbol{r_n}$ from the population Pop and stored in $\boldsymbol{P_Z}$;
Phases of Procreation and Cannibalism
for $\boldsymbol{J = 1}$ to $\boldsymbol{r_n}$ do
    selection of two widows arbitrarily as parents from $\boldsymbol{P_Z}$;
    offspring generation by Eqs. (8) and (9);
    female spiders kill the male spider;
    weaker offspring is destructed by cannibalism rate;
    Store remaining solutions in $\boldsymbol{P_Y}$;
end for
Phases of Mutation
Estimation of mutation number $\boldsymbol{m_n}$ via mutation rate;
for $\boldsymbol{I = 1}$ to $\boldsymbol{m_n}$ do
    Solution selection from $\boldsymbol{P_Z}$;
    generation and mutation of new widows Randomly;
    Store remaining solutions in $\boldsymbol{P_X}$;
end for
Update Process
Update of $\boldsymbol{P = P_Y + P_Z}$;
Attain optimum solution;
Attain optimum solution from $\boldsymbol{P}$;
end
```

3. Experimental Result Analysis

In this section, the performance validation of the proposed model takes place using four videos, collected by our own. The dataset contains four classes namely marriage, politics, school, and college as shown in Figs. 3-4.

Fig. 3. Sample Surveillance Video-College

Fig. 4. Sample Surveillance Video-School

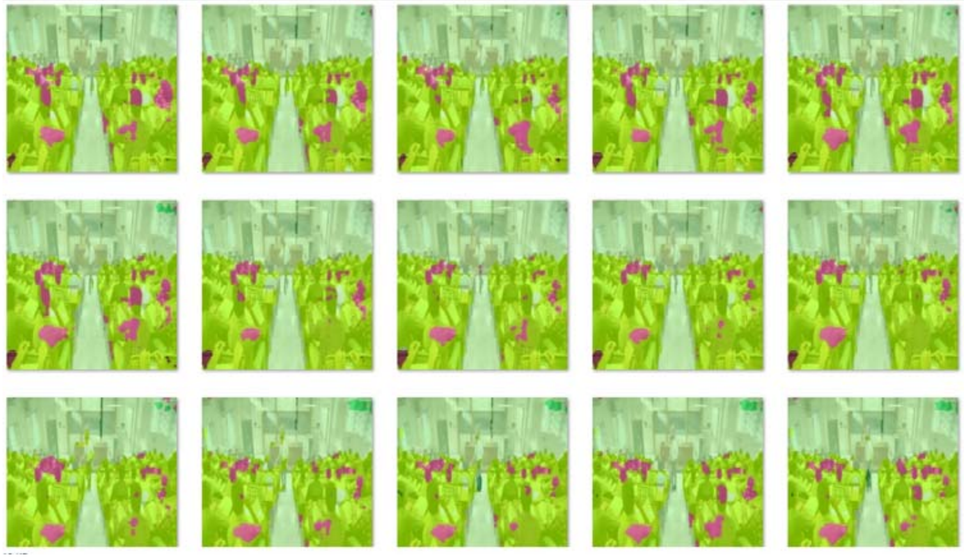

Fig. 5. Results of Deep Instance Segmentation Method

Fig. 5 illustrates the segmented results of the proposed model and the figure highlighted the crowded peoples in the input frames.

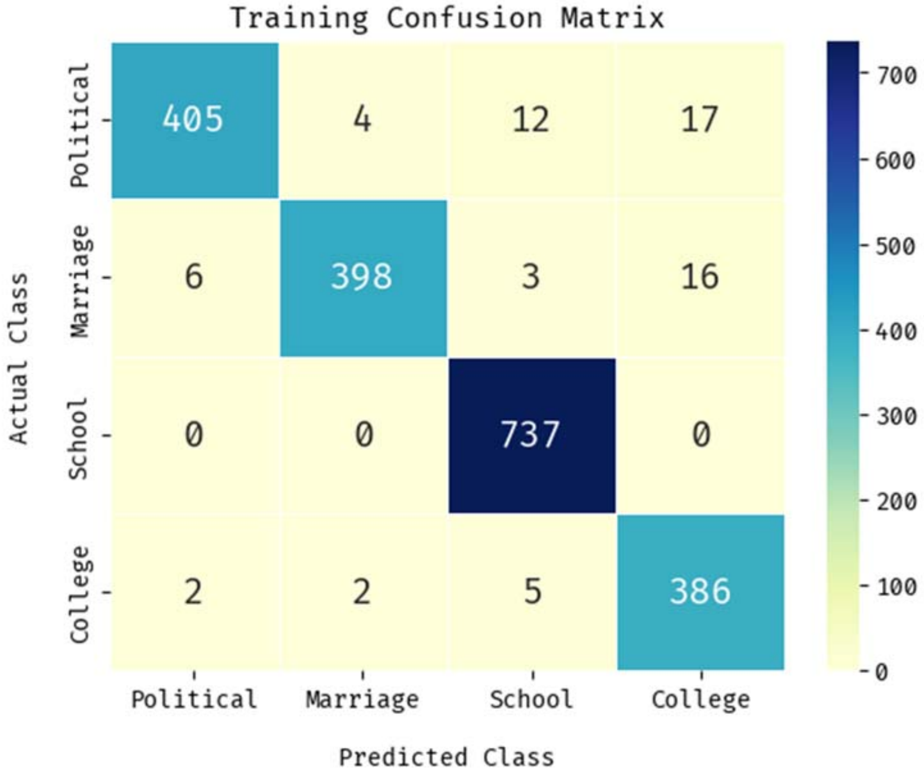

Fig. 6. Confusion matrix analysis of IDL-CBDC technique on training dataset

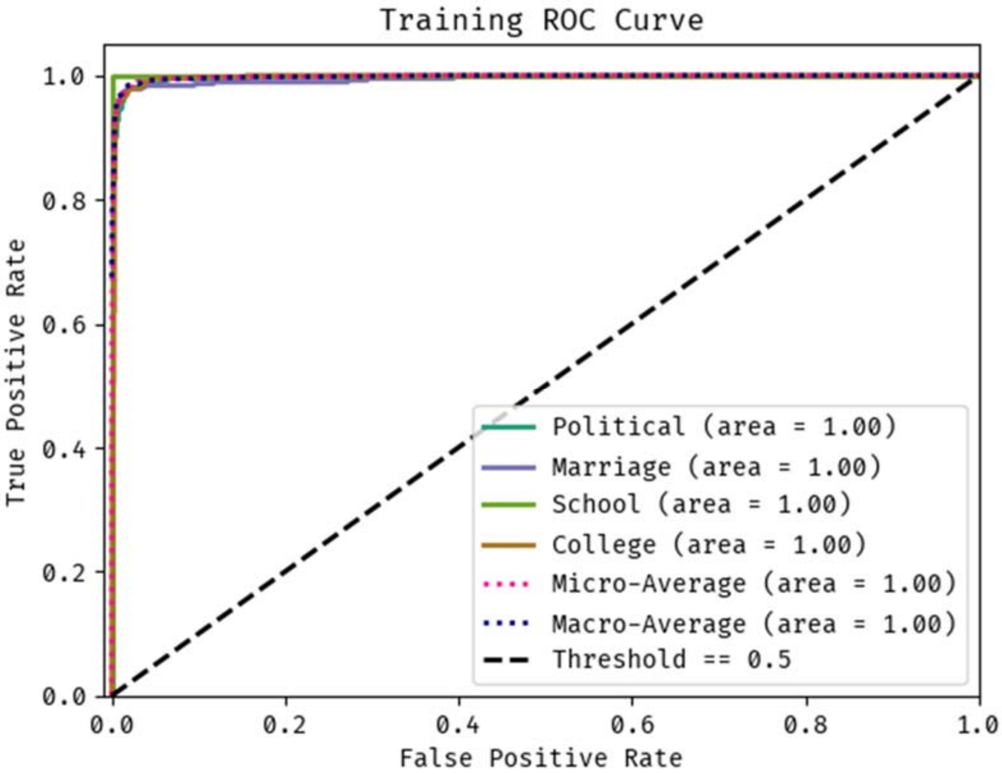

Fig. 7. ROC analysis of IDL-CBDC technique on training dataset

Fig. 6 exhibits the confusion matrix obtained by the proposed model on the training process. The figure showcased that the proposed model has classified 405 instances into political, 398 instances into marriage, 737 instances into school, and 386 instances into college.

The ROC analysis of the proposed model on the training dataset is shown in Fig. 7. The figure reported that the proposed model has resulted in higher ROC of 1.0 under all classes.

Table 1 and Fig. 8 provide the result analysis of the proposed model under different classes. The proposed model has classified the instances under political class with the $accu_y$, $prec\_n$, $sens_y$, $spec_y$and F_{score} of 92.470%, 98.060%, 92.470%, 99.490%, and 95.180% respectively. Likewise, the proposed technique has classified the instances under marriage class with the $accu_y$, $prec\_n$, $sens_y$, $spec_y$and F_{score} of 94.090%, 98.510%, 94.090%, 99.620%, and 96.250% correspondingly. Eventually, the proposed model has classified the instances under school class with the $accu_y$, $prec\_n$, $sens_y$, $spec_y$and F_{score} of 100.000%, 92.120%, 97.720%, 97.930%, and 98.660% correspondingly. Meanwhile, the proposed methodology has classified the instances under College class with the $accu_y$, $prec\_n$, $sens_y$, $spec_y$and F_{score} of 97.720%, 92.120%, 97.720%, 97.930%, and 94.840% correspondingly.

Class	**Accuracy**	**Precision**	**Sensitivity**	**Specificity**	**F1-Score**
Political	92.470	98.060	92.470	99.490	95.180
Marriage	94.090	98.510	94.090	99.620	96.250
School	100.000	97.360	100.000	98.410	98.660
College	97.720	92.120	97.720	97.930	94.840
Average	**96.070**	**96.640**	**96.640**	**98.880**	**96.640**

Table 1. Result analysis of IDL-CBDC technique with distinct measures on training dataset

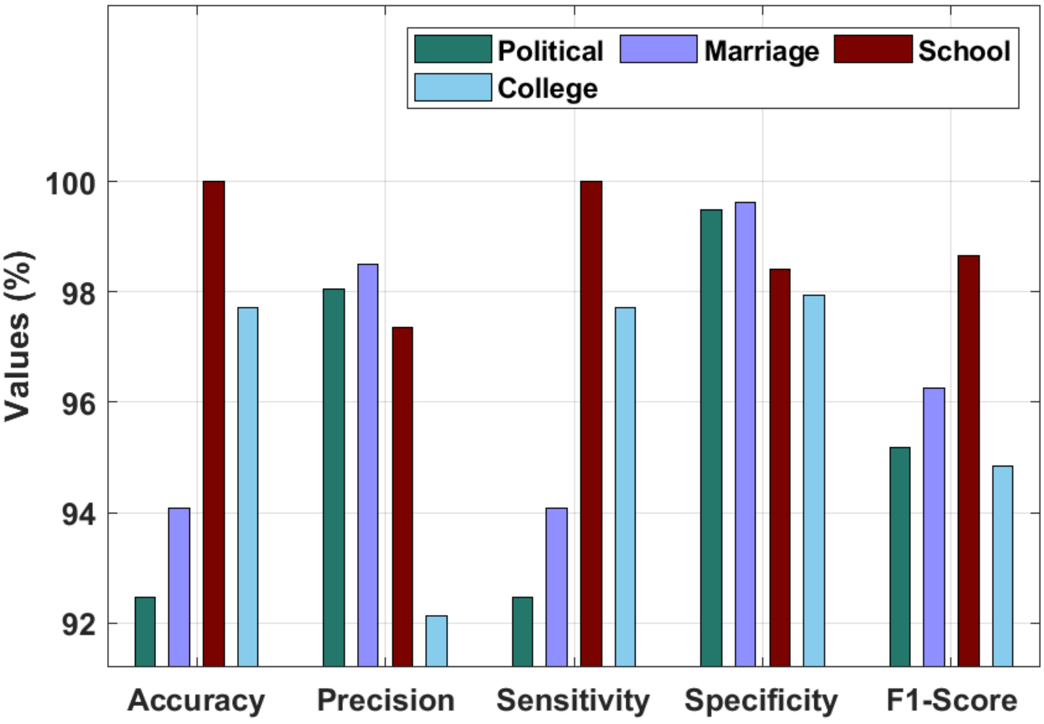

Fig. 8. Result analysis of IDL-CBDC technique on training dataset

Fig. 9 showcases the confusion matrix attained by the proposed method on the testing process. The figure depicted that the proposed approach has classified 172 instances into political, 191 instances into marriage, 304 instances into school, and 166 instances into college.

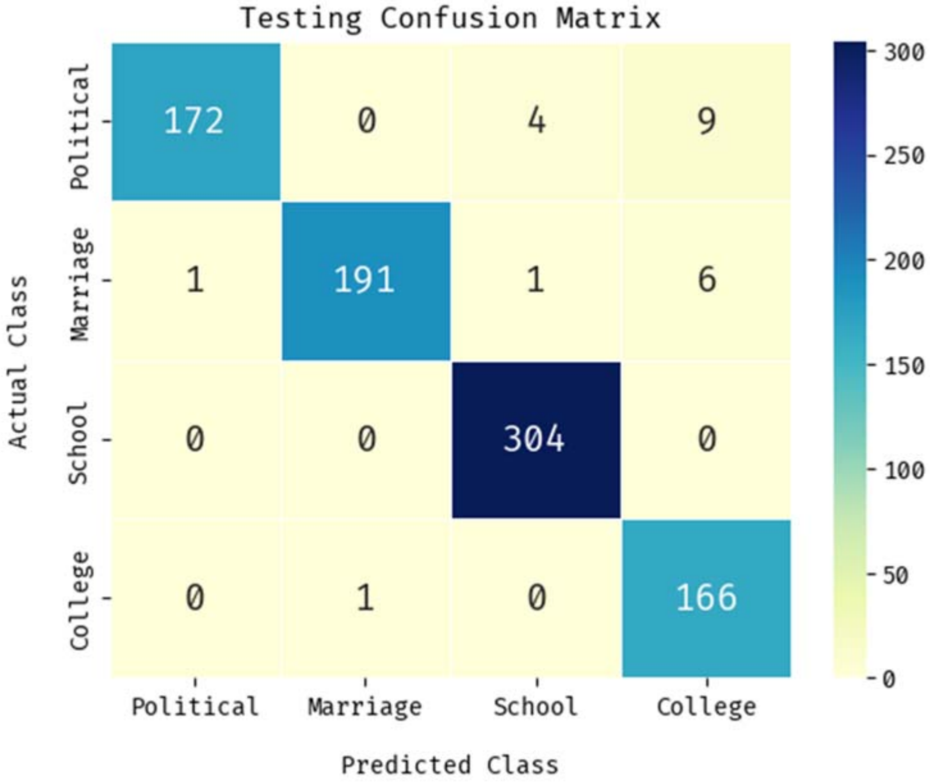

Fig. 9. Confusion matrix analysis of IDL-CBDC technique on testing dataset

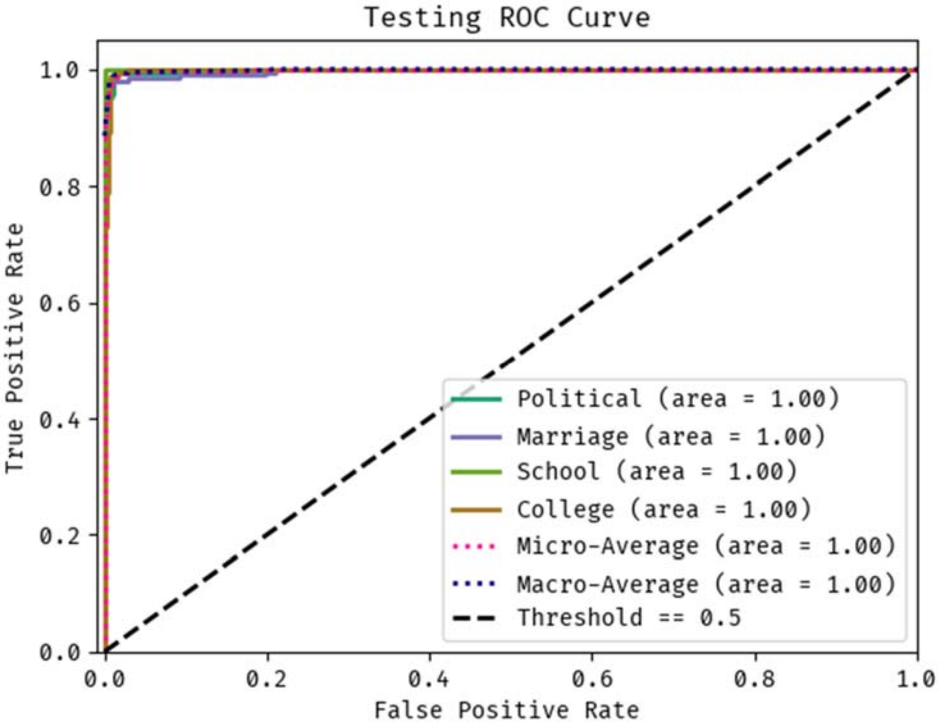

Fig. 10. ROC analysis of IDL-CBDC technique on testing dataset

The ROC analysis of the proposed technique on the testing dataset is illustrated in Fig. 10. The figure reported that the presented technique has resulted to superior ROC of 1.0 under all classes. Table 2 and Fig. 11 offer the outcome analysis of the proposed technique under different classes.

Class	**Accuracy**	**Precision**	**Sensitivity**	**Specificity**	**F1-Score**
Political	92.970	99.420	92.970	99.850	96.090
Marriage	95.980	99.480	95.980	99.850	97.700
School	100.000	98.380	100.000	99.090	99.180
College	99.400	91.710	99.400	97.820	95.400
Average	**97.090**	**97.430**	**97.430**	**99.140**	**97.430**

Table 2. Result analysis of IDL-CBDC technique with distinct measures on testing dataset

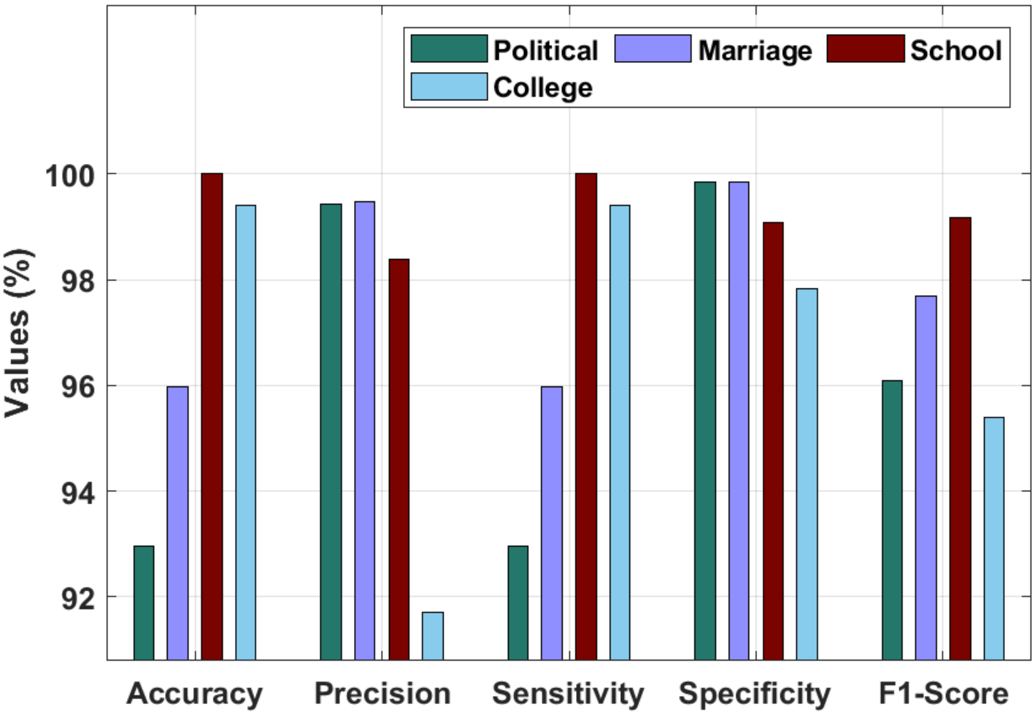

Fig. 11. Result analysis of IDL-CBDC technique on testing dataset

The proposed method has classified the instances under political class with the $accu_y$, $prec\_n$, $sens_y$, $spec_y$and F_{score} of 92.970%, 99.420%, 92.970%, 99.850%, and 96.090% respectively. Likewise, the proposed model has classified the instances under marriage class with the $accu_y$, $prec\_n$, $sens_y$, $spec_y$and F_{score} of 95.980%, 99.480%, 95.980%, 99.850%, and 97.700% correspondingly. Followed by, the projected technique has classified the instances under school class with the $accu_y$, $prec\_n$, $sens_y$, $spec_y$and F_{score} of 100.000%, 98.380%, 100.000%, 99.090%, and 99.180% correspondingly. Finally, the presented technique has classified the instances under College class with the $accu_y$, $prec\_n$, $sens_y$, $spec_y$and F_{score} of 99.400%, 91.710%, 99.400%, 97.820%, and 95.400% correspondingly.

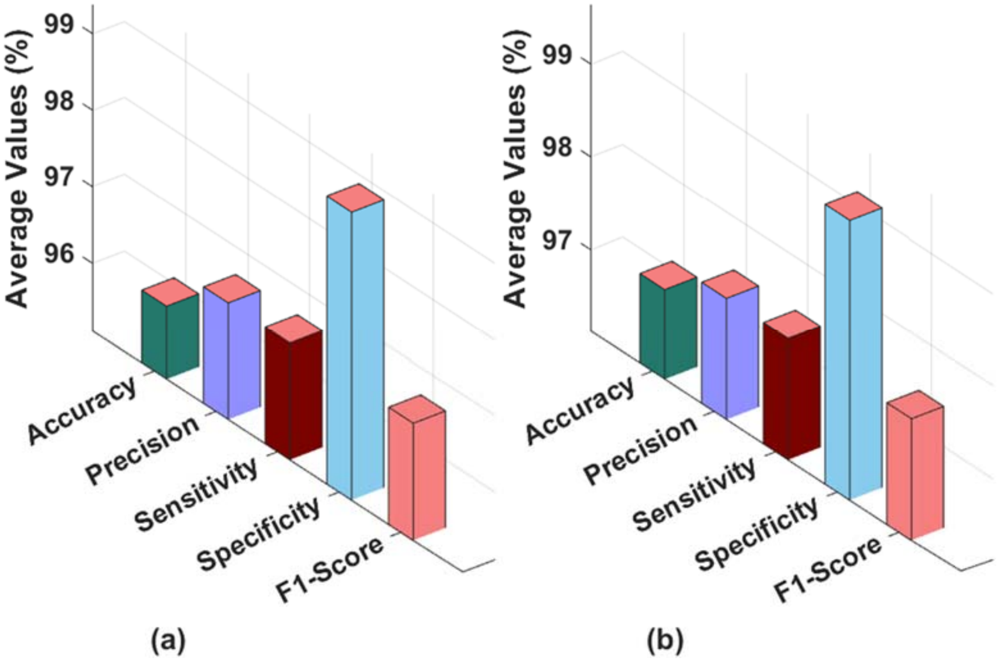

Fig. 12. Average result analysis a) Training Dataset b) Testing Dataset

Fig. 12 shows the average classification results of the proposed model on the test training and testing dataset. On the applied training dataset, the proposed model has resulted in average $accu_y$, $prec\_n$, $sens_y$, $spec_y$and F_{score} of 96.070%, 96.640%, 96.640%, 98.880%, and 96.640% respectively. Likewise, on the applied testing dataset, the proposed technique has resulted in average $accu_y$, $prec\_n$, $sens_y$, $spec_y$and F_{score} of 97.090%, 97.430%, 97.430%, 99.140%, and 97.430% respectively.

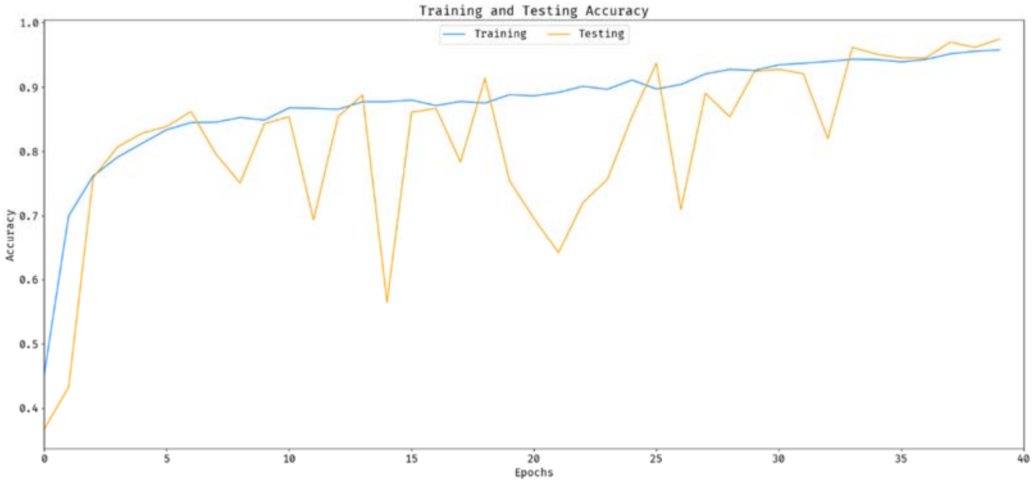

Fig. 13. Accuracy graph analysis of IDL-CBDC technique

Fig. 13 showcases the accuracy analysis of the IDL-CBDC approach on the test dataset. The outcomes depicted that the IDL-CBDC algorithm has accomplished enhanced performance with maximum training and validation accuracy. It can be clear that the IDL-CBDC technique has attained higher validation accuracy over the training accuracy.

Fig. 14 portrays the loss analysis of the IDL-CBDC approach on the test dataset. The results recognized that the IDL-CBDC methodology has resulted in a proficient outcome with lesser training and validation loss. It can be stated that the IDL-CBDC system has offered lower validation loss over the training loss.

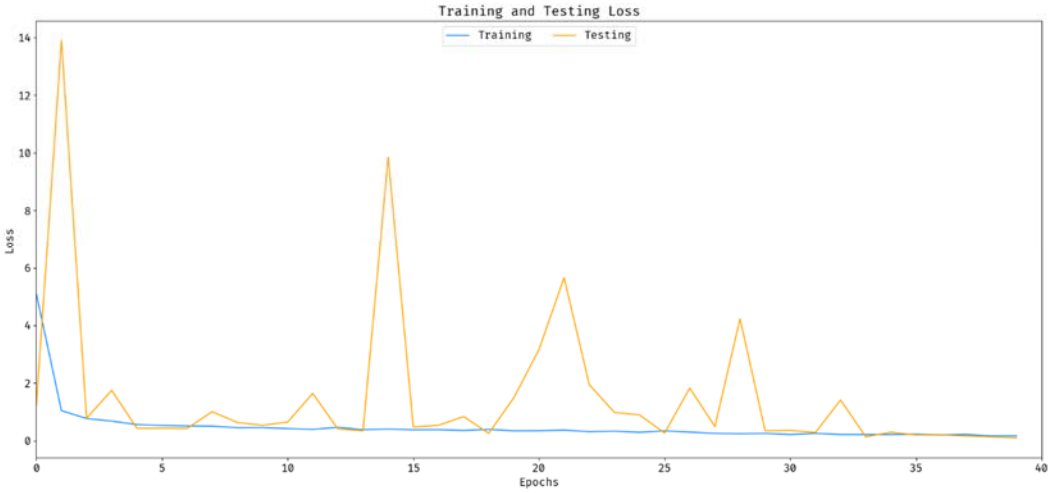

Fig. 14. Loss graph analysis of IDL-CBDC technique

Detailed comparative result analysis of the proposed model with recent approaches takes place in Table 3 [23, 24]. Fig. 15 illustrates the sen_y and $spec_y$ analysis of IDL-CBDC technique with existing approaches. The result depicted that the CD-DLM technique has resulted to lower values of $sens_y$ and $spec_y$. At the same time, the TF-SVM, PS-SVM, M-CNN, and BoW-LBP techniques have obtained moderately closer values of $sens_y$ and $spec_y$. However, the proposed model has resulted in higher $sens_y$ and $spec_y$ of 97.43% and 99.14% respectively.

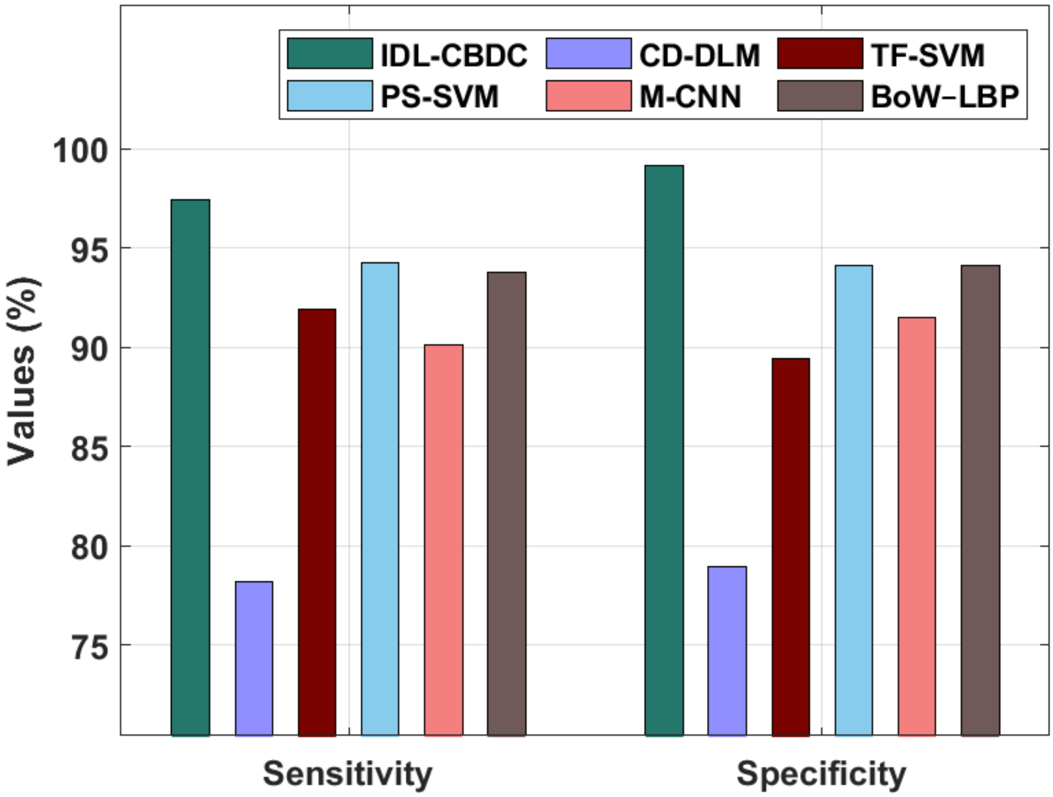

Fig. 15. $Sens_y$ and $Spec_y$ analysis of IDL-CBDC technique

Fig. 16 showcases the $prec_n$ and $F1_{score}$ analysis of IDL-CBDC approach with existing approaches. The outcome demonstrated that the CD-DLM technique has resulted to lower values of $prec_n$ and $F1_{score}$. Likewise, the TF-SVM, PS-SVM, M-CNN, and BoW-LBP methods have gained moderately closer values of $prec_n$ and

$F1_{score}$. At last, the proposed technique has resulted in higher $prec_n$ and $F1_{score}$ of 97.43% and 97.43% respectively.

Methods	**Accuracy**	**Precision**	**Sensitivity**	**Specificity**	**F1-Score**
IDL-CBDC	97.09	97.43	97.43	99.14	97.43
CD-DLM	78.95	77.12	78.19	78.97	77.84
TF-SVM	90.20	91.23	91.94	89.47	89.76
PS-SVM	93.40	95.38	94.27	94.13	93.58
M-CNN	90.67	92.40	90.15	91.48	90.47
BoW–LBP	94.00	94.90	93.80	94.10	94.40

Table 3. Comparative analysis of IDL-CBDC technique with existing algorithms

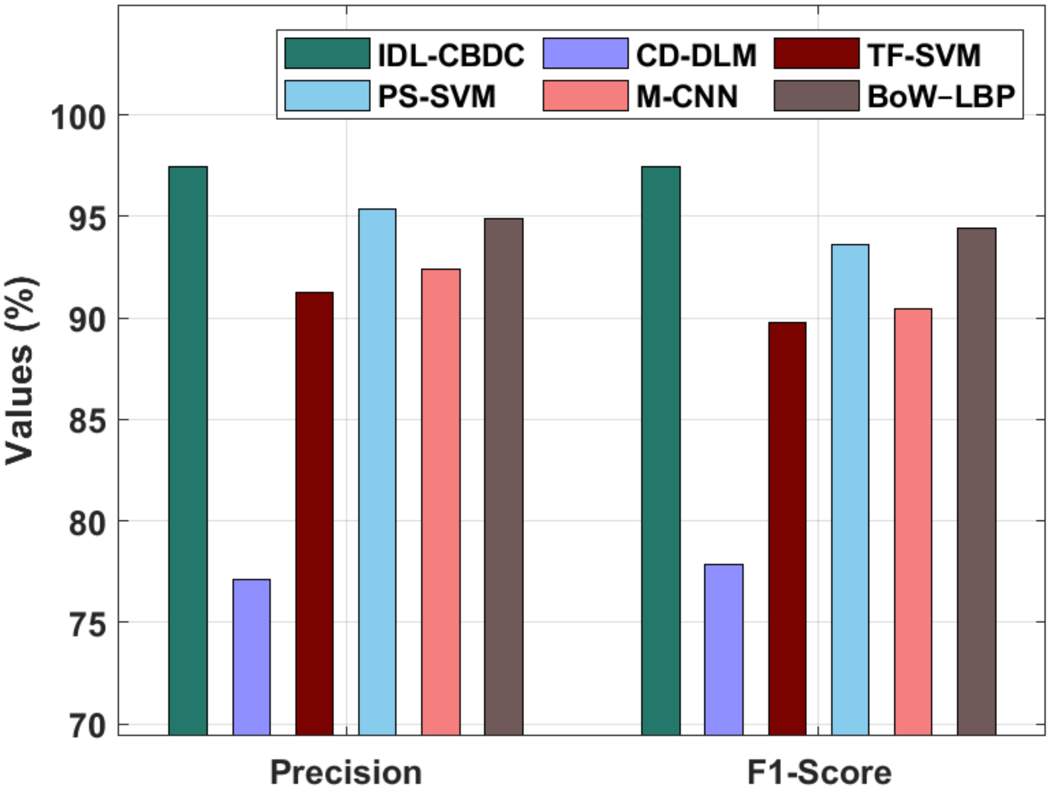

Fig. 16. $Prec_n$ and $F1_{score}$ analysis of IDL-CBDC technique

Fig. 17 depicts the acc_y analysis of IDL-CBDC method with existing algorithms. The outcome outperformed that the CD-DLM system has resulted in least value of acc_y. Followed by, the TF-SVM, PS-SVM, M-CNN, and BoW-LBP techniques have obtained moderately closer values of acc_y. But the proposed methodology has resulted in superior acc_y of 97.09%. Therefore, it is ensured that the proposed model is found to be an effective tool for crowd behavior detection and classification in real time surveillance videos.

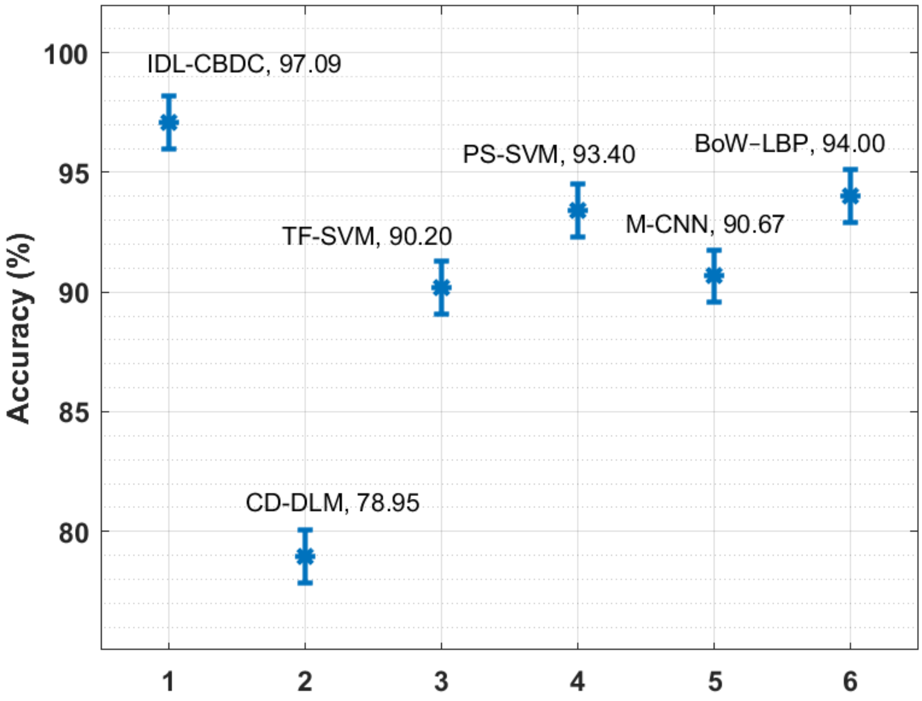

Fig. 17. Acc_y analysis of IDL-CBDC technique with existing approaches

4. Conclusion

In this study, a new IDL-CBDC technique has been derived for crowd behavior analysis in real time surveillance videos, which mainly classifies the video frames into four classes namely marriage, political, school, and college. The proposed IDL-CBDC technique comprises preprocessing, AMF based noise removal, contrast enhancement, DIS based segmentation, ResNet50 based feature extraction, softmax layer based classification, and BWO based hyperparameter optimization. The design of BWO algorithm helps to properly adjust the hyperparameter values such as learning rate, batch size, number of epochs, and number of hidden layers. In order to ensure the improved performance of the IDL-CBDC technique, a set of simulations take place using an own dataset, gathered from public places. Extensive comparative result analysis reported the supremacy of the IDL-CBDC technique over the other techniques. In future, advanced hybrid DL models can be utilized instead of ResNet50 model to improve the classification performance.

Conflicts of Interest: The authors have no conflicts of interest to declare

References

[1] Tursunov, A., Choeh, J.Y. and Kwon, S., 2021. Age and Gender Recognition Using a Convolutional Neural Network with a Specially Designed Multi-Attention Module through Speech Spectrograms. Sensors, 21(17), p.5892.

[2] Abdullah, F., Ghadi, Y.Y., Gochoo, M., Jalal, A. and Kim, K., 2021. Multi-person tracking and crowd behavior detection via particles gradient motion descriptor and improved entropy classifier. Entropy, 23(5), p.628.

[3] Anvarjon, T.; Kwon, S.; Mustaqeem. Deep-net: A lightweight CNN-based speech emotion recognition system using deep frequency features. Sensors 2020, 20, 5212.

[4] Zhan, B.; Monekosso, D.; Remagnino, P.; Velastin, S.; Xu, L.Q. Crowd analysis: A survey. Mach.Vis. Appl. 2008, 19, 345–357.

[5] Idrees, H.; Soomro, K.; Shah, M. Detecting humans in dense crowds using locally-consistent scale prior and global occlusion reasoning. IEEE Trans. Pattern Anal. Mach. Intell. 2015, 37, 1986–1998.

[6] Yücesoy, E.; Nabiyev, V.V.J.C.; Engineering, E. A new approach with score-level fusion for the classification of a speaker age and gender. Comput. Electr. Eng. 2016, 53, 29–39.

[7] Landge, M.B.; Deshmukh, R.; Shrishrimal, P.P. Analysis of variations in speech in different age groups using prosody technique. Int. J. Comput. Appl. 2015, 126, 14–17.

[8] Rodriguez, M.; Laptev, I.; Sivic, J.; Audibert, J.Y. Density-aware person detection and tracking in crowds. In Proceedings of the 2011 IEEE International Conference on Computer Vision (ICCV), Barcelona, Spain, 6–13 November 2011; pp. 2423–2430.

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

[10] Herrmann, C.; Metzler, J. Density estimation in aerial images of large crowds for automatic people counting. Proc. SPIE 2013, 8713, 87130V

[11] Pawar, K. and Attar, V., 2019. Deep learning approaches for video-based anomalous activity detection. World Wide Web, 22(2), pp.571-601.

[12] Rezaei, F. and Yazdi, M., 2021. Real-time crowd behavior recognition in surveillance videos based on deep learning methods. Journal of Real-Time Image Processing, pp.1-11.

[13] Varghese, E.B. and Thampi, S.M., 2018, August. A deep learning approach to predict crowd behavior based on emotion. In International Conference on Smart Multimedia (pp. 296-307). Springer, Cham.

[14] Gao, M., Jiang, J., Ma, L., Zhou, S., Zou, G., Pan, J. and Liu, Z., 2019, June. Violent crowd behavior detection using deep learning and compressive sensing. In 2019 Chinese Control And Decision Conference (CCDC) (pp. 5329-5333). IEEE.

[15] Song, B. and Sheng, R., 2020. Crowd Counting and Abnormal Behavior Detection via Multiscale GAN Network Combined with Deep Optical Flow. Mathematical Problems in Engineering, 2020.

[16] Shukla, H. S., Narendra Kumar, and R. P. Tripathi. "Median Filter based Wavelet Transform for Multilevel Noise." International Journal of Computer Applications 107.14 (2014): 11-14.

[17] Gao, Z., 2018, December. An Adaptive Median Filtering of Salt and Pepper Noise based on Local Pixel Distribution. In the 2018 International Conference on Transportation & Logistics, Information & Communication, Smart City (TLICSC 2018).

[18] Zhao H, Shi J, Qi X, Wang X, Jia J. Pyramid Scene Parsing Network. Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition. 2017 July 21-26; Honolulu, HI, United states. New York: IEEE 2017.

[19] Zhang, Z., Gao, S. and Huang, Z., 2021. An automatic glioma segmentation system using a multilevel attention pyramid scene parsing network. Current Medical Imaging, 17(6), pp.751-761.

[20] Chu, Y., Yue, X., Yu, L., Sergei, M. and Wang, Z., 2020. Automatic image captioning based on ResNet50 and LSTM with soft attention. Wireless Communications and Mobile Computing, 2020.

[21] Hayyolalam V, Kazem AAP (2020) Black widow optimization algorithm: a novel meta-heuristic approach for solving engineering optimization problems. Eng Appl Artif Intell 87:103249.

[22] Ravikumar, S. and Kavitha, D., 2021. IOT based autonomous car driver scheme based on ANFIS and black widow optimization. Journal of Ambient Intelligence and Humanized Computing, pp.1-14.

[23] Jia, D., Zhang, C. and Zhang, B., 2021. Crowd density classification method based on pixels and texture features. Machine Vision and Applications, 32(2), pp.1-22.

[24] Meynberg, O., Cui, S. and Reinartz, P., 2016. Detection of high-density crowds in aerial images using texture classification. Remote Sensing, 8(6), p.470.

Authors Profile

Mr. Sivachandiran.S received his M. Tech degree in Computer Science and Engineering from SRM University, Chennai. He is currently pursuing a Ph. D from Annamalai University, Tamil Nadu. His area of research is Image processing, Deep learning, IoT, and edge computing.

Dr.K.Jagan Mohan received his Ph.D. in Computer Science and Engineering from Annamalai University, Tamilnadu. Currently, he is working as an Associate Professor in the Department of Information Technology, Annamalai University, Tamil Nadu. He has more than 20 years of teaching and research experience. He is currently involved in research work on Image Processing, Deep Learning, Computer Networks, and IoT.

Dr. Mohammed Nazer received his Ph.D. in Computer Science from Prist University, Tamilnadu. Currently, he is working as Professor in the Department of Computer Science and Principal, Raak Arts and Science College, Tamil Nadu. He has more than 20 years of teaching and research experience. He is currently involved in research work on Image Processing, Network security, and IoT.