IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.Doi Number

Design of Artificial Intelligence Driven Crowd Density Analysis for Sustainable Smart Cities

Shtwai Alsubai[1], Ashit Kumar Dutta[2], Faisal Alghayadh[3], Bader Hussain Alamer[4], Radha Mohan Pattanayak[5,*], Janjhyam Venkata Naga Ramesh[6], Sachi Nandan Mohanty[7]

[1]Department of Computer Science, College of Computer Engineering and Sciences in Al-Kharj, Prince Sattam bin Abdulaziz University, P.O. Box 151, Al-Kharj 11942, Saudi Arabia; drShtwaiAlsubai@gmail.com

[2]Department of Computer Science and Information Systems, College of Applied Sciences, AlMaarefa University, Ad Diriyah, Riyadh, 13713, Kingdom of Saudi Arabia; adotta@um.edu.sa

[3]Department of Computer Science and Information Systems, College of Applied Sciences, AlMaarefa University, Ad Diriyah, Riyadh, 13713, Kingdom of Saudi Arabia; fghayadh@um.edu.sa

[4]Department of Emergency Medical Services, College of Applied Sciences, AlMaarefa University, Diriyah, 13713, Riyadh, Saudi Arabia; grafi@um.edu.sa

[5]School of Computer Science and Engineering(SCOPE), VIT-AP University, Amaravati, Andhra Pradesh, India; radhamohan.pattanayak@gmail.com

[6]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur Dist., Andhra Pradesh - 522302, India; jvnramesh@gmail.com

[7]School of Computer Science & Engineering (SCOPE), VIT -AP University, Andhra Pradesh, India; Sachinandan09@gmail.com

Corresponding author: Radha Mohan Pattanayak (e-mail: pattanayak@gmail.com).

ABSTRACT Smart Cities refer to urban areas which exploits recent technologies for improving the performance, sustainability, and livability of their infrastructure and services. Crowd Density Analysis (CDA), a vital component of Smart Cities, involves the use of sensors, cameras, and data analytics to monitor and analyze the density and movement of people in public spaces. CDA utilizing DL harnesses the control of neural networks to mechanically and exactly evaluate the density of crowds in numerous settings, mainly in smart cities. DL techniques like Recurrent Neural Network (RNN) and Convolutional Neural Network (CNN), are trained on vast datasets of crowd videos or images to learn complex designs and features. These models can forecast crowd density levels, recognize crowd anomalies, and offer real-time visions into crowd behavior. This study designs an Artificial Intelligence Driven Crowd Density Analysis for Sustainable Smart Cities (AICDA-SSC) technique. The aim of the AICDA-SSC method is to analyze the crowd density and classify it into multiple classes by the use of hyperparameter-tuned DL models. To accomplish this, the AICDA-SSC technique applies contrast enhancement using the CLAHE approach. Besides, the complex and intrinsic features can be derived by the use of the Inception v3 model and its hyperparameters can be chosen by the use of the marine predator's algorithm (MPA). For crowd density detection and classification, the AICDA-SSC technique applies a gated recurrent unit (GRU) model. Finally, a chaotic sooty tern optimizer algorithm (CSTOA) based hyperparameter selection procedure takes place to increase the effectiveness of the GRU system. The experimental evaluation of the AICDA-SSC technique takes place on a crowd-density image dataset. The experimentation values showcase the superior performance of the AICDA-SSC method to the recently developed DL models.

INDEX TERMS Crowd Density; Artificial Intelligence; Sustainable; Chaotic Sooty Tern Optimization; Smart Cities

I. INTRODUCTION

A smart city is a full entity that combines tangible and non-tangible sources for use by its people [1]. The compatibility of ICT-based infrastructures with fixed infrastructures is essential to become a system efficient in providing predictable services to people, without disturbing or negotiating the excellence of life [2]. It is also considered an advanced design to provide sustainable development and service by satisfying the 6 sizes of sustainability namely environment, people, economy, mobility, governance, and living. As a crucial part and one of the 6 sizes for developing smart cities, mobility must be given special and significant attention [3]. The concept of supportable transport methods as an essential part

of emerging smart cities is being beaten. Because the conventional methods of transport assumed in traditional cities are joined with in-built tasks like congestion issues, accident occurrence, pollution, and much more, these are dangers to the method and delay the socio-economic actions of smart cities [4]. Urban surveillance has changed from grainy CCTV footage to high-definition streams managed by neural networks. These networks are trained on huge datasets that can differentiate among the movement of vehicles and people, deduce crowd creations, and even recognize unusual designs that might direct possible dangers [5].

Over the past years, crowd analysis (CA) has revealed steady development owing to the arrival of new techniques [6]. Deep learning (DL) methods have been increasingly utilized for numerous uses owing to discriminatory control and effectual functional removal. Many models employed in traditional CA were improper for modern surveillance due to definite restrictions. Normally, modern surveillance methods are categorized by intense worries and dynamicity in crowd motion styles and the functioning situations of surveillance tools [7]. This dissimilar feature can confuse the usage of present models in the analysis and monitoring of the dense crowd. CA researchers must improve new models to reply to the worry in the novel atmosphere where computer vision (CV) is gradually required to observe and analyze numerous people from video footage of the surveillance cameras in real-time [8]. This involves evaluating the assortment of the crowd and the density distribution across the complete collection area. Recognizing regions above security can aid in delivering prior alarms and could stop crowd crushes [9]. The estimation of the number of crowds also aids in measuring the significance, logistics, and substructure of the event. With the improvement of hardware technology and the DL model, the performance of numerous CV tasks has been significantly enhanced, and CNN has played a significant part in many tasks namely image classification, target recognition, and semantic segmentation [10]. So, CNN was commonly employed in calculating tasks, and the related performance has been enhanced.

This study designs an Artificial Intelligence Driven Crowd Density Analysis for Sustainable Smart Cities (AICDA-SSC) technique. To achieve this, the AICDA-SSC model uses contrast enhancement using the CLAHE approach. Besides, the complex and intrinsic features can be derived by the use of the Inception v3 model and its hyperparameters can be chosen by the use of the marine predators' algorithm (MPA). For crowd density detection and classification, the AICDA-SSC technique applies a gated recurrent unit (GRU) model. Finally, a chaotic sooty tern optimizer algorithm (CSTOA) based hyperparameter selection procedure takes place to improve the efficacy of the GRU system. The experimental evaluation of the AICDA-SSC technique takes place on crowd crowd-density image dataset. In short, the key contributions of the paper are listed as follows.

- An automated AICDA-SSC technique comprising Inception v3 feature extractor, MPA-based hyperparameter tuning, GRU classification, and CSTOA based parameter selection has been developed for crowd density detection. In order to the best of our knowledge, the AICDA-SSC technique never existed in the literature.
- Combines the Inception v3 structure as feature extraction, improving the model's capability to take and examine difficult patterns within crowd density data, thus refining the accuracy of crowd analysis in smart cities.
- Uses the MPA to modify hyperparameters, enhancing the performance of the crowd density study method. MPA's bio-inspired optimizer technique donates to the efficacy and flexibility of the model.
- Includes GRU as portion of the AICDA-SSC method, allowing the technique to efficiently take and study time-based dependences in crowd density information, improving the accuracy of forecasts over time.
- Executes the CSTOA for hyperparameter range, donating to the sturdiness and flexibility of the technique by dynamically fine-tuning parameters dependent upon chaotic optimizer principles.
- The incorporation of MPA and CSTOA reproduces a hybrid optimization approach, merging bio-inspired models and chaotic optimizer principles to improve the model's flexibility and performance.

II. RELATED WORKS

Zhu et al. [11] develop a comprehensive AI-based CA model structure for rail transport stations, by examining and picturing CA information from video frames of highest density crowds. Then a general AI helped organizational structure (AI Crowd) was developed. Deep SORT and YOLO have been combined into the model structure. Camera calibration is employed to convert identified paths into a real-life organized method. Bai et al. [12] reflect on crowd mindset and other aspects and begin a fixed basic technique of crowd assembly designs. To fuse actual multi-granularity surveillance videos with dissimilar viewpoints, multi-column CNNs (M-CNNs) have been employed to remove the local compactness features of the crowd in a lower-altitude viewpoint.

In [13], the KSA in crowd organization utilizing AI through the Hajj was tested to generate a technique for the same situations. This research used the descriptive systematic model. The Arc Gis Pro 2.9.2 is utilized to generate maps. A strategic study was also directed to the KSA in crowd organization utilizing a SWOT study. Rezaee et al. [14] inspect UAV excess and irregular population action designs. Furthermore, the aim is to examine accepted video frames from UAVs. Yu et al. [15] developed a result of actual 3D visualization of outside acts permitted by Visual IoT (V-IoT) and AI. Primary, LiDAR, infrared cameras, gas sensors,

airborne and fixed cameras, etc. are used to gather multi-modal information and forward it to the clouds. AI techniques have been implemented in the cloud. Lastly, the clouds guide the AI model outcomes to the visual method in station devices.

Solmaz et al. [16] developed a novel adaptive machine learning (ML) method, termed CountMeIn, to find the crowd estimate issue by employing neural networks and polynomial regression. CountMeIn adjusts Wi-Fi utilizing the proficient method and preserves the highest accuracy after training for an extended period without cameras. Ahmed et al. [17] proposed an IoT-based crowd surveillance method that utilizes a DL model. Dual virtual lines are definite to total how many people are going or arriving at the part. Prezioso et al. [18] project a new framework that integrates the general YOLO object recognition system with innovative CA methods. The developed structure influences YOLO's actual object recognition abilities to discover numerous things within video frames and concentrate on finding people.

Al Duhayyim et al. [19] provide Aquila Optimizer with TL based Crowd Density Analysis for Sustainable Smart Cities (AOTL-CDA3S) method. This model aims to recognize dissimilar types of crowd densities in the smart cities. In [20], a new system design is defined for real-time crowd detection. Also, a privacy-aware platform that assists the application of AI devices using identified Wi-Fi traces is also suggested. Guastella et al. [21] offer the HybridIoT model over an estimate model that assimilates heterogeneous data attained from a few dissimilar sensors. Yang et al. [22] presents a novel technique. Initially, a crowd density estimate model dependent upon Tencent user density (TUD) data is constructed for dissimilar times in open public spaces.

Padmaja et al. [23] presented work to make a method that can categorize normal and abnormal crowd Behaviour utilizing an actual time video surveillance method to identify abnormalities and observe crowded metropolitan zones. Zhou et al. [24] proposes MJPNet-S*, a multistyle joint-perception network for RGB-T/D crowd density estimation on drones. This model incorporated a novel trimodal module and a two-step hybrid model, together with a lightweight student network assisted by neighboring collaborative distillation. In [25], the E2C model, together with systematic evaluation and intellectual computing models is proposed, for carbon emissions in visual computing.

III. THE PROPOSED MODEL

In this work, an AICDA-SSC method is presented. The purpose of the AICDA-SSC method is to analyze the crowd density and classify it into multiple classes by the use of hyperparameter-tuned DL models. To accomplish this, the AICDA-SSC technique aims to identify different kinds of processes such as preprocessing, feature extractor, classification, and parameter tuning process. Fig. 1 shows the complete procedure of the AICDA-SSC system.

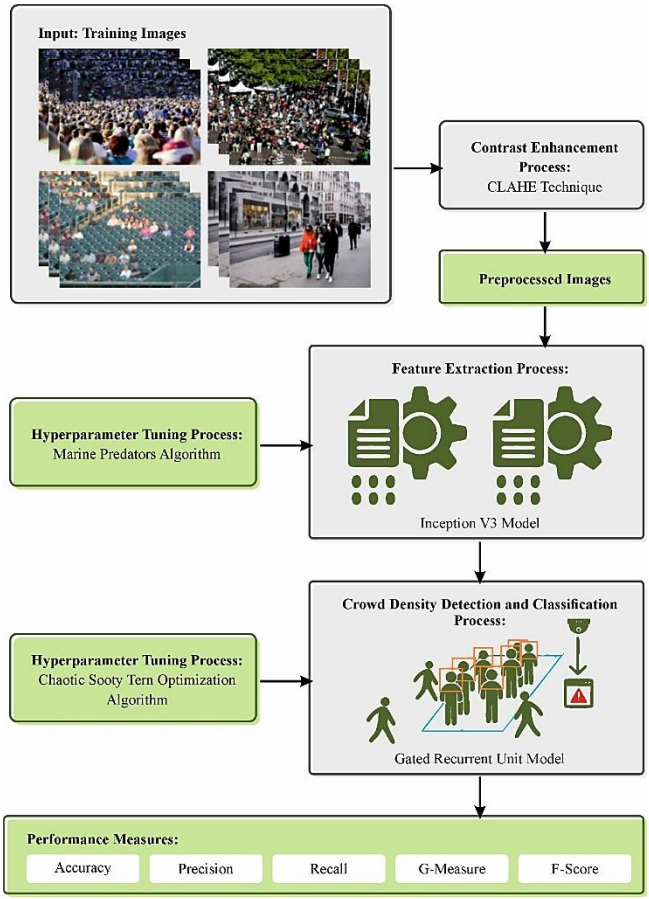

FIGURE 1. Overall procedure of the AICDA-SSC technique

A. PREPROCESSING

At the primary level, the AICDA-SSC technique applies contrast enhancement using the CLAHE approach. Resizing and contrast enhancement processes are performed during the preprocessing stage [26]. The initial stage is to resize the input images and the CLAHE approach is used to improve the difference. The CLAHE technique improves image clarity with the lowest contrast. It has shown to be a robust mechanism for optimizing and organizing digital images. CLAHE includes dual vital advantages beyond the histogram equalization technique. Initially, the CLAHE model is predictable to improve the intensity of individual pixels more evenly. By using the classical histogram equalization technique, the array of the histogram is extended, which results in a more even distribution of grey values throughout the image.

CLAHE improves the difference in every zone, resulting in an enhanced contrast throughout the whole image.

$$D_B = \frac{255}{8 \times 8} \sum_{i=0}^{D_A} H(i) \quad (1)$$

Next, CLAHE can reduce the problem of noise enhancement by restricting contrast enhancement. The histogram height is enlarged to L to improve the visual symbol while retaining the similar histogram as follows:

$$H(i) == \begin{cases} H(i) + L, & H(i) < H_{\max} \\ HH(i)_{max_{max}} \end{cases} \quad (2)$$

B. INCEPTION V3 MODEL

For the feature extraction process, the Inception v3 can be employed. The CNN could remove data features layers by layers over the slip function of the convolutional kernel and has been widely used in various fields [27]. The main way to increase the performance of a system is by enlarging the network depth and width, but it results in difficult network training and overfitting.

The Inception model is a powerful tool to resolve the problem, which exploits convolutional kernel of dissimilar sizes for similar layer feature maps, then passed over kernel size 1×1 convolution for the reduction of channel dimensionality, lastly, the channel splicing is summarized for the extraction of feature data. While expanding the network width, the amount of parameters remains equal. InceptionV3 is a deep CNN architecture that has made important contributions to the area of CV and image recognition. Presented by Google researchers, it signifies a prominent development in DL techniques. InceptionV3 uses an exclusive and new inception unit, integrating equivalent convolutions of dissimilar dimensions within the similar layer, permitting the system to capture multi-scale features competently. This architecture enables the removal of complex hierarchical designs from images, making it mainly effectual for tasks such as object detection and image classification. InceptionV3 is well-known for its extraordinary accuracy on large-scale image databases, like ImageNet, and its capability to simplify well to varied visual detection tasks. Its victory lies in arresting a balance between model difficulty and computational efficacy, making it a general choice for numerous CV applications in either industry or research.

Batch normalization is commonly applied between the activation function and the convolutional layer, which standardizes the data in the channel size, which could efficiently improve training speed, resolve the gradient disappearing problem, and decrease the over-fitting phenomenon. The mathematical expression is:

$$\mu_B = \frac{1}{m}\sum_{i=1}^{m} x_i \quad (3)$$

$$\sigma_B^2 = \frac{1}{m}\sum_{i=1}^{m} (x_i - \mu_B)^2 \quad (4)$$

$$\widehat{x_i} = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 - \varepsilon}} \quad (5)$$

$$y_i = \gamma \widehat{x_i} + \beta \quad (6)$$

Where the average of the training batch is represented as μ_B, the training batch variance is σ_B^2, the data in a single network is m, and the i^{th} data on the feature maps before and after normalize, pan and zoom are $x_i, \widehat{x_i}$, and y_i, correspondingly. β and γ denote the translation and spatial scaling of control factors, correspondingly; the constant ε is evaluated when the variance is 0.

Next, the hyperparameters can be chosen by the use of MPA. The MPA replicas the foraging behavior strategy of marine predators [28]. The Levy fly in a lower prey attention atmosphere and Brownian effort in sufficient regions. The metaheuristic presents an equilibrium among Brownian as well as Levy tactics for dissimilar movement features of solution candidates. This optimizer procedure is separated into 3 stages, approaching the relative speeds among predator and prey, and exhibiting the Fish Aggregation Device (FAD) effect reliable for the hunter's behavioral variations. Besides, since predators regularly arrive in previously visited and foraging regions, the procedure remembers them.

In an initialized phase of this population-based system, all individual solutions have been evenly spread as follows:

$$X_0 = X_{\min} + r(md(X_{\max} - X_{\min}), \quad (7)$$

Where X_{max} and $X_{\min}$ denote the upper and lower limits and $rand$ is an even random vector in $[0 \text{ and } 1]$. The procedure rules the performance of dual kinds of individuals signified by Elite and Prey in Eqs. (8) and (9) matrixes, respectively.

$$Elite = \begin{bmatrix} X_{1,1}^I & X_{1,2}^I & \cdots & X_{1,d}^I \\ X_{2,1}^I & X_{2,2}^I & \cdots & X_{2,d}^I \\ \cdots & \cdots & \cdots & \cdots \\ X_{n,1}^I & X_{n,2}^I & \cdots & X_{n,d}^I \end{bmatrix}. \quad (8)$$

From the abovementioned equation, d represents the size of the problem, X^1 signifies the best predator vector and n refers to the amount of individual's solution.

$$Prey = \begin{bmatrix} X_{1,1} & X_{1,2} & \cdots & X_{1,d} \\ X_{2,1} & X_{2,2} & \cdots & X_{2,d} \\ \cdots & \cdots & \cdots & \cdots \\ X_{n,1} & X_{n,2} & \cdots & X_{n,d} \end{bmatrix}. \quad (9)$$

In the equation of Prey, the jth dimension of the ith prey is represented by $X_{i,j}$.

The stages in the MPA are dependent upon the development of the optimizer procedure. For the 1st one-third of iteration, the procedure arrives at Stage 1, where the highest velocity ratio is imitated, i.e., the hunter moves quicker than the target. Therefore, the ith prey, $i = 1,2, \ldots,n$ is defined in Eq. (10).

$$Prey_i = Prey_i + P * R \otimes \big(R_B \otimes (Elite_i - R_B \otimes Prey_i)\big), \quad (10)$$

Where P refers to the constant number $(P = 0.5)$, R_B denotes the vector of generally dispersed normal numbers, R signifies the vector of evenly spread random numbers in $[0 \text{ and } 1]$, and$'\otimes'$ represents the elementwise multiplication. In Stage 1, the stage of exploration plays the main part, while, in the next one-third of iterations, exploitation and exploration stages have been similarly arranged. Here, the algorithm arrives at Stage 2, pretending the unit velocity ratio. At this stage, the ith prey has been upgraded using its location in the population.

$$Prey_i = \begin{cases} Prey_i + P * R \otimes \left(R_L \otimes (Elite_i - R_L \otimes Prey_i)\right), & if\ i \le n/2 \\ Elite_i + P * CF \otimes \left(R_B \otimes (R_B \otimes Elite_i - Prey_i)\right), & otherwise, \end{cases} \quad (11)$$

Where R_L denotes the random vector numbers and $CF = \left(1 - \frac{iter}{iter_{\max}}\right)^{\frac{2*iter}{irer_{\max}}}$. The CF is employed to handle the stage size of the hunter's effort. In Stage 3, which survives to the finale of the optimizer procedure, the lowest velocity percentage is reflected. Where the ith prey is measured as exposed in Eq. (12).

$$Prey_i = Elite_i + P * CF \otimes \left(R_L \otimes (R_L \otimes Elite_i - Prey_i)\right) \quad (12)$$

The marine predator's memory is pretended by the fitness contrast of the present candidates from the preceding iterations, declining worst candidates.

To simulate the effect of FADs, the algorithm adjusts the Prey, i.e.

$$Prey_i = \begin{cases} Prey_i + CF[X_{\min} + R \otimes (X_{\max} - X_{\min})] \otimes U, & if\ r \le FADs \\ Prey_i + [FADs(1-r) + r](Prey_{r1} - Prey), & otherwise, \end{cases} \quad (13)$$

Whereas FADs are the effect probability of (0.2), U is the dual vector generated by a threshold vector in the range of [0 and 1]. The binarization adapts the values higher than 0.2 to 1 and sets the residual to 0. The $r1$ and $r2$ have been chosen preys at random, $r1, r2 \in \{1,2, \dots, n\}$.

C. CROWD DENSITY CLASSIFICATION USING GRU

The AICDA-SSC technique uses the GRU technique for crowd density recognition and classification. This part describes the common procedure of GRU. It is an innovative form of Standard RNN [29]. The LSTM contains 3 gates which will not preserve the interior cell layer but are combined into the hidden layer (HL) of the RNN. This data has been moved to the next GRU. Numerous gates of GRU are definite below.

1) UPDATE GATE

It describes several prior knowledge sent to the prospect. It defines the equivalent output gate in the LSTM recurrent unit. It has been expressed utilizing

$$y = \sigma\left(W^{(y)} Z_t + V^{(y)} H_{t-1}\right) \quad (14)$$

where Z_t denotes the unit of the network that is multiplied by the value of the weight $W^{(y)}$. That is forwarded to the HL H_{t-1}, which contains the data of the prior layers and is increased by its weight values $V^{(y)}$. These dual grades have been included to offer the last outcome in the update gate among 0 and 1. This can be signified many past data needs to be sent to the prospect. It can be employed to remove the danger of gradient issues.

2) RESET GATE

It signifies how much earlier information can be neglected. It matches an input gate and ignores it in an LSTM recurrent unit.

$$R = \sigma\left(W^{(y)} Z_t + V^{(y)} H_{t-1}\right) \quad (15)$$

The dual outcomes are included and then increased by their values of weight, and the task of sigmoid value is functional to the output solutions.

3) CURRENT MEMORY GATE

This type of gate is combined into the reset gate. It delivers the non-linearity input and offers 0 mean input. It is employed in order to decrease the result that prior data of the existing data is sent to the prospect gate. The current memory gate calculation is executed utilizing

$$H = \tanh\left(W z_t + R \odot V H_{t-1}\right) \quad (16)$$

The input z_t and HL are increased by their value of weight. Then, calculate the product of Hadamard among the reset gates $R \odot V H_{t-1}$. Next, enlarge the initial output and second procedure values. The tanh nonlinear activation function has been used to analyze the present memory content. Lastly, compute the HL values that contains the existing values and direct it to the system, executed by the update gate. It defines the present memory content and is expressed by utilizing

$$H_t = y_t \odot H_{t-1} + (1 - y_t) \odot H_t \quad (17)$$

Implement elementwise multiplication to the update gate and compute to $(1 - y_t) \odot H_t$. Lastly, add these dual outcomes for computing the present memory content of the GRU. It improves the RNN memory capability and resolves the vanishing gradient issues.

The main improvement of GRUs will be the usage of gating devices to handle the flow of data inside and outside of the cell. The gating device contains 2 gates such as reset and update. The projected bi-directional RNN contains dual layers of RNN which can procedure simultaneously Y is input and Y_k signifies the dissimilar inputs with diverse time stamps. The processing has been executed simultaneously, and the layers are organized consecutively. The HL of RNN contains dual HLs defined for each timestep. The HLs are joined into one layer by including dual inputs utilizing an easy addition function. Bi-directional RNN performs every neuron of the system. The feature match procedure reflects the output and input features to process the timestep. RNN contains Soft GRU which has a low complexity, and procedures the input $\tilde{x}_{k,n}$ utilizing past $h_{k-1,n}$. This procedure has low difficulty owing to the soft plus and activation function. The developed Soft GRU is definite as.

$$h_{k,n} = (1 - z_t) \odot h_{k-1,n} + z_t \odot \tilde{x}_{k,n} \quad (18)$$

$$\tilde{x}_{k,n} = \pi\left(W_x x_{k,n} + b_x\right) \quad (19)$$

$$z_t = \sigma\left(W_x x_{k,n} + U_z x_{k-1,n} + b_z\right) \quad (20)$$

Whereas $\pi(x)$ signifies the soft plus function that is calculated as $(1 + e^x)$ and σ signifies the function of the sigmoid and b_X, b_z denotes the biased values, and W_X, W_z, and U_z epitomizes the weight values. GRU can work for the natural language process, which uses both historical and present information about traffic, social media, weather, and roads that are accessible in the edge servers.

D. CSTOA-BASED PARAMETER TUNING

Finally, the CSTOA-based hyperparameter selection process takes place to improve the efficiency of the GRU model. STOA is a novel search optimizer technique dependent upon the feeding behaviors of sooty terns that comprises two stages: a global search stage which pretends terns migrate and

a local search stage where the tern circles and attacks the target [30]. Like other population techniques, every sooty tern signifies a search agent, and each search agent-organized constructs the matrix X. During initialization stage, the method begins by creating an early matrix within the search range.

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_i \\ \vdots \\ X_N \end{bmatrix} = \begin{bmatrix} x_{11} & \cdots & x_{1j} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{i1} & \cdots & x_{ij} & \cdots & x_{im} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{Nj} & \cdots & x_{Nm} \end{bmatrix}_{N\times m} \quad (21)$$

In Eq. (4), the population matrix is X; the value of the search agent at a size is x_{ij}; the vector of the search agent, representing an initial random solution at the initial phase is X_i and updated at the iterative computation, the index of the search agent is i; the size index is j; the size of search space is m; and the number of populations is N.

1) MIGRATION BEHAVIOR (GLOBAL SEARCH)

Sooty terns (ST) are involved in the behavior of migrants in the hunt for plentiful food resources. The main reason for this stage is to rapidly recognize the optimum region by searching for the overall search stage at random [31]. This stage contains 3 parts namely conflict avoidance, update of position, and aggregation.

2) CONFLICT AVOID

To evade crashes among individuals throughout the movement, extra mass S_A has been developed in the iterative computation to upgrade single locations.

$$C = S_A \times X(t) \quad (22)$$

Whereas C refers to the location in the instance of no collisions with other individuals, $X(t)$ denotes the search agent position, t signifies the present iteration index, S_A denotes an extra variable employed to evade collisions that are intended as:

$$S_A = c_f \times \left(1 - \frac{t}{T}\right) \quad (23)$$

Where c_f denotes the constant utilized to alter S_A, which is normally fixed to 2, T refers to the highest amount of iterations. So, S_A will consecutively reduce from two to $zero$.

3) AGGREGATION

After evading collisions among adjacent agents, the search agents will travel near the finest location amid the adjacent agents, that is, near the location of the optimum result which is stated below:

$$M = C_B \times \left(X_{besi}(t) - X(t)\right) \quad (24)$$

Here, M signifies the procedure of traveling X at dissimilar positions near the X_{est} place of the optimum solutions, and C_B states the random number utilized for making the search more complete. The expression is as below:

$$C_B = 0.5 \times Rand \quad (25)$$

Where Rand denotes the random value among $(0 and 1)$.

4) UPDATE OF POSITION

The search agent upgrades its location dependent upon the finest location, which is expressed as:

$$D = C + M \quad (26)$$

Here, D is the space among the present individuals and the global optimum positions.

5) BEHAVIOR OF MIGRATION (LOCAL SEARCH)

If ST wants to affect their target at the time of migration, suddenly they will fly in a curved form in the sky. The equation has been stated below:

$$\begin{cases} x' = R \times \sin(i) \\ y' = R \times \cos(i) \\ z' = R \times i \\ R = u \times e^{kv} \end{cases} \quad (27)$$

Where u and v denote the constants that define the form of the spirals and have a value of 1. i denotes the random variable that exists amid the range of $(0 \; and \; 2\pi)$. R represents the radius of every spiral. The position update of search agent formulation is as:

$$X(t) = (D \times (x' + y' + z \times X_{besi}(t) \quad (28)$$

The spatial distribution of the initialized solution set in the metaheuristic algorithm is a crucial factor influencing the optimization result of the algorithm and global search speed. Based on the random number strategy, the early solution set produced in the STOA is hard to evenly issue within the search range which leads to a reduction in the search efficacy. Chaotic mapping has the features of regularity, randomness, and traversal are applied for initializing the position of STOA individuals employing chaotic sequence to prevent falling into local extreme. For the initial population, Circle chaotic mapping is applied. Circle mapping generates chaotic sequence expression, as follows:

$$num_{i+1} = mod\left(num_i + 0.2 - \frac{0.5}{2\pi} \times \sin(2\pi \times num_i), 1\right) \quad (29)$$

In Eq. (29), num_i indicates the value of i^{th} chaotic sequence and mod represents the residual operation.

Assume$[Z_{\min}, Z_{\max}]$ as the search range of optimization objective problem and the chaotic sequence value of the circle map is num_{ij}, the solution vector X is formulated by Eq. (30):

$$x_{ij} = Z_{\min} + (z_{\max} - Z_{\min}) \times num_{ij} \quad (30)$$

Algorithm 1: Pseudocode of STOA
Input: the population size N, and Population initialization X randomly. Output: Optimal searching agent, Xbest 1: begin STOA 2: Parameter initialization of SA and CB 3: Determine the fitness of every search agent 4: Xbest ← optimal searching agent 5: While (t<T) do 6: for every searching agent do 7: Update the position of searching agent 8: end for 9: Update the variables SA and CB 10: Determine the fitness of every search agent 11: Update Xbest if there is a better solution than the former optimum solution 12: t ← t+1

```
13:     end while
14: return Xbest
15: end
```

The CSTOA model originates a fitness function (FF) to reach boosted classification effectiveness. It defines a positive integer to represent the superior execution of the candidate outcome. In this paper, the error rate of classifier reductions is examined as FF, as displayed in Eq. (31).

$$fitness(x_i) = ClassifierErrorRate(x_i) = \frac{No.of\ misclassified\ instances}{Total\ no.of\ instances} \times 100 \quad (31)$$

IV. PERFORMANCE VALIDATION

This part examines the crowd density recognition results of the AICDA-SSC technique on the dataset comprising four classes and 1000 samples as definite in Table 1. Fig. 2 determines the sample pictures.

TABLE I
DETAILS OF THE DATASET

Labels	Class Names	No. of Instances
Class0	Dense Crowd	250
Class1	Medium Dense Crowd	250
Class2	Sparse Crowd	250
Class3	No Crowd	250
Total No. of Instances		1000

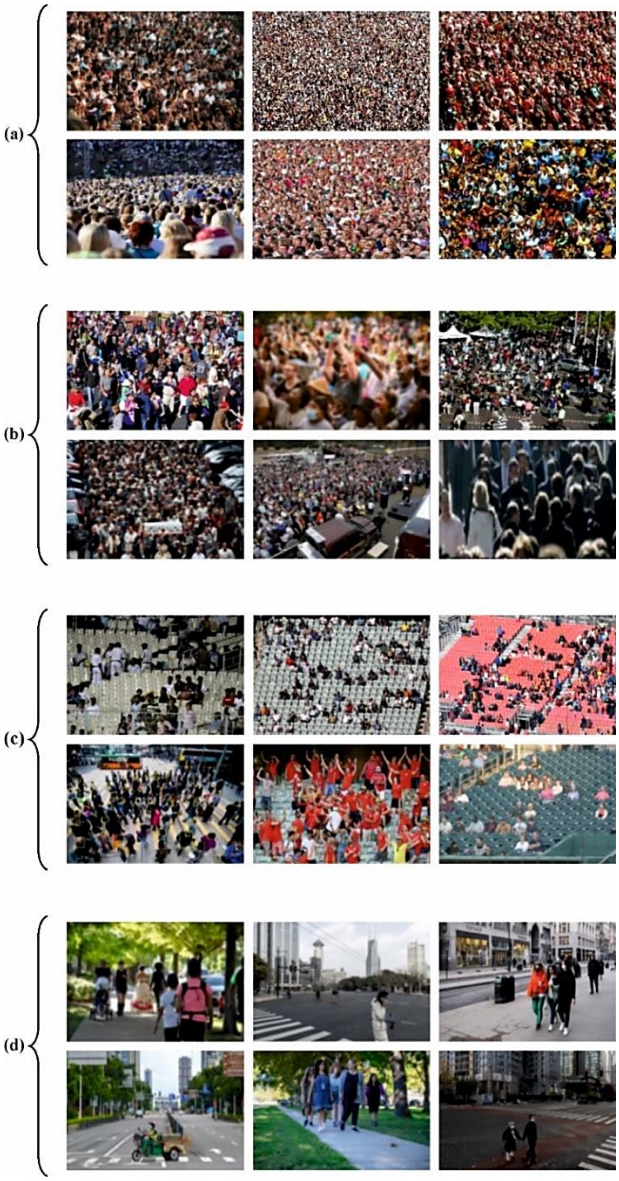

FIGURE 2. a) Dense Crowd b) Medium Dense Crowd c) Sparse Crowd d) No Crowd

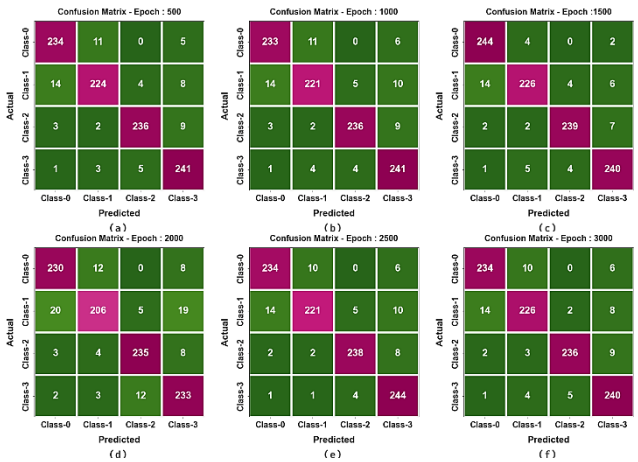

FIGURE 3. Confusion matrices of AICDA-SSC technique (a-f) Epochs 500-3000

Fig. 3 exhibits the confusion matrices created by the AICDA-SSC model below numerous epochs. The outcomes specify the effective recognition and classification of four classes properly.

The crowd density recognition outcome of the AICDA-SSC technique can be examined under varying epochs in Table 2 and Fig. 4. These experimentation outcomes indicate that the AICDA-SSC model accurately recognizes four classes below all epochs. With 500 epochs, the AICDA-SSC technique attains an average $accu_y$ of 96.75%, $prec_n$ of 93.54%, $reca_l$ of 93.50%, F_{score} of 93.49%, and $G_{measure}$ of 93.51%. In addition, with 1000 epochs, the AICDA-SSC model gains an average $accu_y$ of 96.55%, $prec_n$ of 93.15%, $reca_l$ of 93.10%, F_{score} of 93.09%, and $G_{measure}$ of 93.11%.

TABLE II
CROWD DENSITY RECOGNITION OUTCOME OF AICDA-SSC TECHNIQUE UNDER SEVERAL EPOCHS

Class Labels	$Accu_y$	$Prec_n$	$Reca_l$	F_{Score}	$G_{Measure}$
Epoch500					
Class-0	96.60	92.86	93.60	93.23	93.23
Class-1	95.80	93.33	89.60	91.43	91.45
Class-2	97.70	96.33	94.40	95.35	95.36
Class-3	96.90	91.63	96.40	93.96	93.99
Average	96.75	93.54	93.50	93.49	93.51
Epoch1000					
Class-0	96.50	92.83	93.20	93.01	93.01
Class-1	95.40	92.86	88.40	90.57	90.60
Class-2	97.70	96.33	94.40	95.35	95.36
Class-3	96.60	90.60	96.40	93.41	93.46
Average	96.55	93.15	93.10	93.09	93.11
Epoch1500					
Class-0	97.70	93.49	97.60	95.50	95.52
Class-1	96.50	95.36	90.40	92.81	92.85
Class-2	98.10	96.76	95.60	96.18	96.18
Class-3	97.50	94.12	96.00	95.05	95.05
Average	97.45	94.93	94.90	94.88	94.90
Epoch2000					
Class-0	95.50	90.20	92.00	91.09	91.09
Class-1	93.70	91.56	82.40	86.74	86.86
Class-2	96.80	93.25	94.00	93.63	93.63
Class-3	94.80	86.94	93.20	89.96	90.02
Average	95.20	90.49	90.40	90.35	90.40
Epoch2500					
Class-0	96.70	93.23	93.60	93.41	93.41
Class-1	95.80	94.44	88.40	91.32	91.37
Class-2	97.90	96.36	95.20	95.77	95.78
Class-3	97.00	91.04	97.60	94.21	94.27
Average	96.85	93.77	93.70	93.68	93.71
Epoch3000					
Class-0	96.70	93.23	93.60	93.41	93.41
Class-1	95.90	93.00	90.40	91.68	91.69
Class-2	97.90	97.12	94.40	95.74	95.75
Class-3	96.70	91.25	96.00	93.57	93.60
Average	96.80	93.65	93.60	93.60	93.61

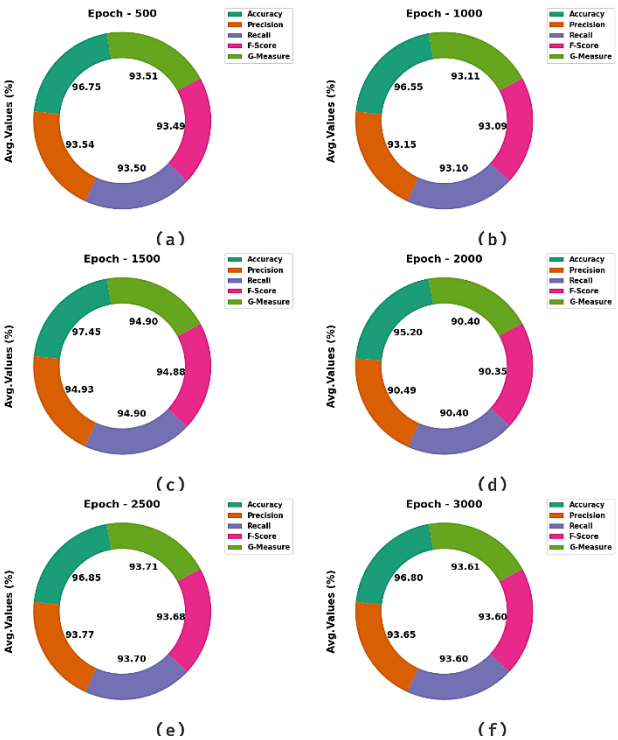

FIGURE 4. Average of AICDA-SSC technique (a-f) Epochs 500-3000

Followed by, with 1500 epochs, the AICDA-SSC method gains an average $accu_y$ of 97.45%, $prec_n$ of 94.93%, $reca_l$ of 94.90%, F_{score} of 94.88%, and $G_{measure}$ of 94.90%. Moreover, with 2500 epochs, the AICDA-SSC algorithm attains an average $accu_y$ of 96.85%, $prec_n$ of 93.77%, $reca_l$ of 93.70%, F_{score} of 93.68%, and $G_{measure}$ of 93.71%. Lastly, with 3000 epochs, the AICDA-SSC system gains an average $accu_y$ of 96.80%, $prec_n$ of 93.65%, $reca_l$ of 93.60%, F_{score} of 93.60%, and $G_{measure}$ of 93.61%.

The $accu_y$ curves for training (TRA) and validation (VL) exposed in Fig. 5 for the AICDA-SSC algorithm at epoch 1500 deliver valued visions into its performance. Mainly, there is a steady development in both TRA and TES $accu_y$ to increasing epochs, demonstrating the model's capability to absorb and recognize designs from both TRA and TES data. The upward trend in TES $accu_y$ underlines the model's adaptability to the TRA dataset and its ability to generate accurate estimates on hidden data, highlighting strong generalized capabilities.

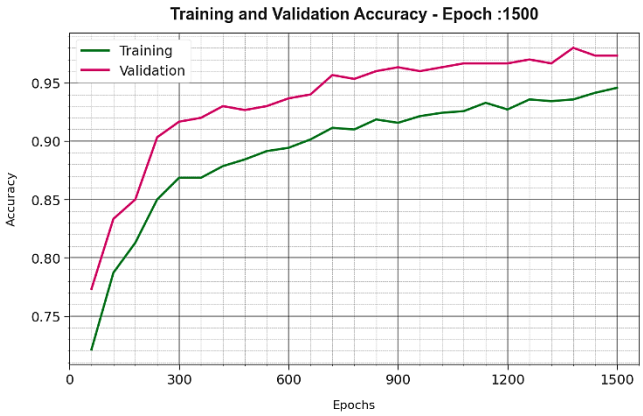

FIGURE 5. $Accu_y$ **curve of AICDA-SSC technique under epoch 1500**

Fig. 6 examines a comprehensive outline of the TRA and TES loss values for the AICDA-SSC technique under epoch 1500. The TRA loss gradually minimization as the model improves its weights to diminish classification errors. The loss curves exhibit the model's location with the TRA data, emphasizing its capability to arrest designs efficiently. Notable is the incessant modification of parameters in the AICDA-SSC method, projected to reduce variances amongst forecasts and real TRA labels.

FIGURE 6. **Loss curve of AICDA-SSC technique under epoch 1500**

Regarding the precision-recall (PR) curve presented in Fig. 7, the results approve that the AICDA-SSC technique under epoch 1500 steadily attains upgraded PR values across each class. These results highlight the model's actual ability to discriminate among dissimilar classes, highlighting its efficiency in precisely diagnosing class labels.

Furthermore, in Fig. 8, we existing ROC curves formed by the AICDA-SSC technique under epoch 1500, representing its ability to differentiate among classes. These curves provide valuable visions into how the trade-off between TPR and FPR differs through diverse classification epochs and thresholds. The outcomes emphasize the model's precise classification efficiency under numerous class labels, underlining its efficacy to overcome various classification processes.

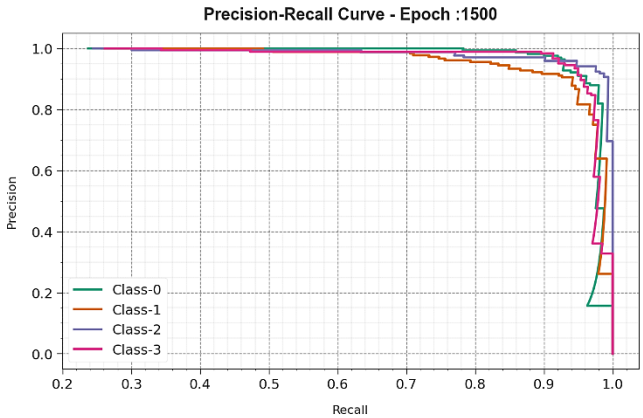

FIGURE 7. **PR curve of AICDA-SSC technique under epoch 1500**

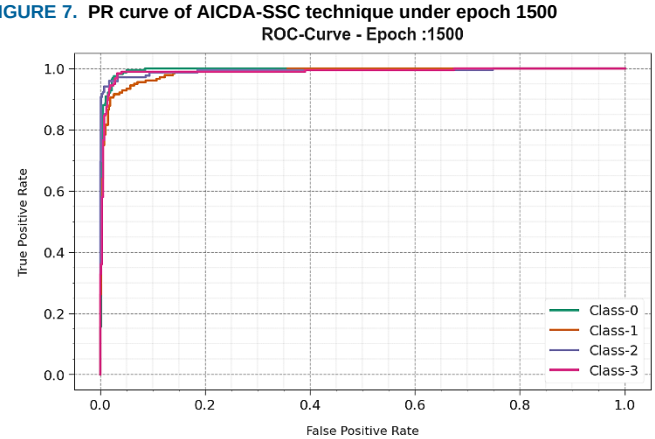

FIGURE 8. **ROC curve of AICDA-SSC technique under epoch 1500**

For guaranteeing the promising results of the AICDA-SSC model, a wide-ranging comparison assessment is made in Table 3 [32]. Fig. 9 examines a comparative $prec_n$ and $reca_l$ study of the AICDA-SSC method. The results indicate that the Gabor and BoW-SRP model has presented the lowest values of $prec_n$ and $reca_l$. Next, the BoW-LBP and GLCM-SVM techniques have obtained slightly boosted values of $prec_n$ and $reca_l$. Followed by, the GoogleNet and VGGNet models have demonstrated moderately closer values of $prec_n$ and $reca_l$. Although the MDTL-ICDDC model reaches near-optimal $prec_n$ and $reca_l$ of 92.90% and 92.90%, the AICDA-SSC technique demonstrates maximum $prec_n$ and $reca_l$ values of 94.93% and 94.90%, respectively.

TABLE III
COMPARATIVE OUTCOME OF AICDA-SSC TECHNIQUE WITH OTHER APPROACHES

Methods	$Prec_n$	$Reca_l$	$Accu_y$	F_{Score}
Gabor	61.83	62.30	71.83	61.98
BoW-SRP	68.33	67.85	80.40	67.88
Bow-LBP	74.68	74.15	84.04	74.35
GLCM-SVM	75.47	73.52	79.78	87.99
GoogleNet	82.98	85.26	84.40	81.00
VGGNet	86.14	82.78	84.75	84.95
MDTL-ICDDC	92.90	92.90	96.45	92.87

AICDA-SSC	94.93	94.90	97.45	94.88

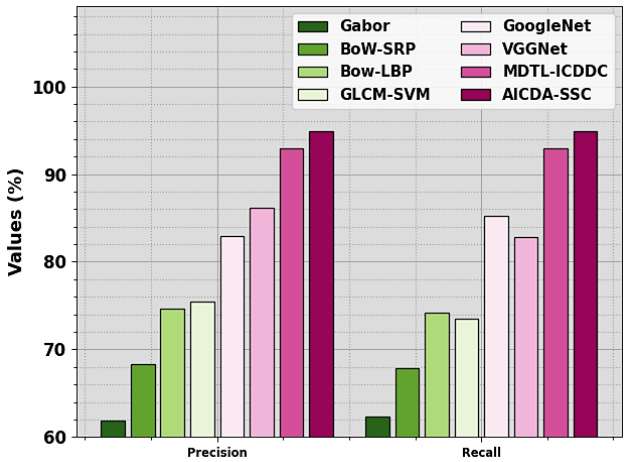

FIGURE 9. $Prec_n$ **and** $reca_l$ **the outcome of the AICDA-SSC technique with other approaches**

Fig. 10 surveys a comparative $accu_y$ and F_{score} study of the AICDA-SSC method. The results specify that the Gabor and BoW-SRP techniques have demonstrated minimum values of $accu_y$ and F_{score}. Then, the BoW-LBP and GLCM-SVM algorithms have attained moderately enhanced values of $accu_y$ and F_{score} . Next, the GoogleNet and VGGNet systems have verified moderately nearer values of $accu_y$ and F_{score} . Although the MDTL-ICDDC system attains near optimum $accu_y$ and F_{score} of 96.45% and 92.87%, the AICDA-SSC methodology determines the largest $accu_y$ and F_{score} values of 97.45% and 94.88%, correspondingly.

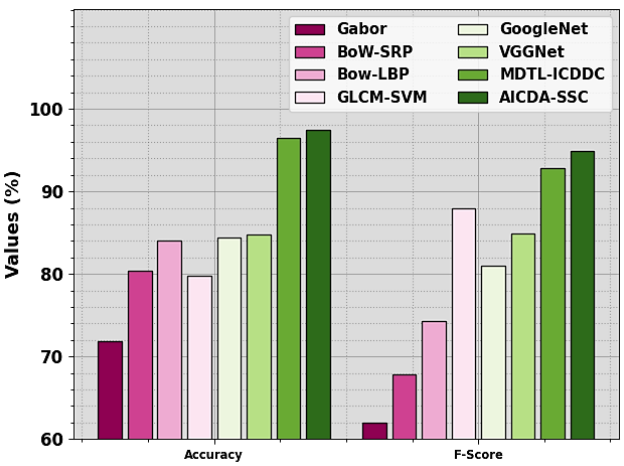

FIGURE 10. $Accu_y$ **and** F_{score} **the outcome of the AICDA-SSC technique with other approaches**

These outcomes highlighted the greater performance of the AICDA-SSC algorithm in the crowd density recognition procedure.

V. CONCLUSION

In this manuscript, an AICDA-SSC method is developed. The purpose of the AICDA-SSC technique is to analyze the crowd density and classify it into multiple classes by the use of hyperparameter-tuned DL models. To accomplish this, the AICDA-SSC technique applies contrast enhancement using the CLAHE technique. Besides, the complex and intrinsic features can be derived by the use of the Inception v3 model and its hyperparameters can be chosen by the use of MPA. For crowd density detection and classification, the AICDA-SSC technique applies the GRU model. Finally, a CSTOA-based hyperparameter selection process take place to increase the efficacy of the GRU model. The experimental evaluation of the AICDA-SSC approach takes place on crowd crowd-density image dataset. The obtained values of the AICDA-SSC approach showcase the greater accuracy outcome of 97.45% over recently developed DL models. The AICDA-SSC technique face restrictions in adaptability and scalability to dynamic urban atmospheres. Future study may concentrate on improving scalability and robustness, and also to explore adaptive mechanisms for accommodating growing urban dynamics and diverse crowd characteristics.

REFERENCES

[1] Z. Fan, H. Zhang, Z. Zhang, G. Lu, Y. Zhang, and Y. Wang, 'A survey of crowd counting and density estimation based on convolutional neural network', *Neurocomputing*, vol. 472, pp. 224–251, Feb. 2022, doi: 10.1016/j.neucom.2021.02.103.

[2] D. Garcia-Retuerta, P. Chamoso, G. Hernández, A. S. R. Guzmán, T. Yigitcanlar, and J. M. Corchado, 'An Efficient Management Platform for Developing Smart Cities: Solution for Real-Time and Future Crowd Detection', *Electronics*, vol. 10, no. 7, p. 765, Mar. 2021, doi: 10.3390/electronics10070765.

[3] G. Sreenu and M. A. Saleem Durai, 'Intelligent video surveillance: a review through deep learning techniques for crowd analysis', J Big Data, vol. 6, no. 1, p. 48, Dec. 2019, doi: 10.1186/s40537-019-0212-5.

[4] G. Gao, J. Gao, Q. Liu, Q. Wang, Y. Wang, "Cnn-based density estimation and crowd counting: A survey." arXiv 2020, arXiv:2003.12783, doi : https://doi.org/10.48550/arXiv.2003.12783

[5] H. Fradi and J.-L. Dugelay, "Towards crowd density-aware video surveillance applications", *Information Fusion*, vol. 24, pp. 3–15, Jul. 2015, doi: 10.1016/j.inffus.2014.09.005.

[6] G. Solmaz, P. Baranwal and F. Cirillo, "CountMeIn: Adaptive Crowd Estimation with Wi-Fi in Smart Cities," *2022 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, Pisa, Italy, 2022, pp. 187-196, doi: 10.1109/PerCom53586.2022.9762354.

[7] H. Minoura, R. Yonetani, M. Nishimura and Y. Ushiku, "Crowd Density Forecasting by Modeling Patch-Based Dynamics," in *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 287-294, April 2021, doi: 10.1109/LRA.2020.3043169.

[8] A. Fitwi, Y. Chen, H. Sun, R. Harrod, "Estimating interpersonal distance and crowd density with a single-edge camera". *Computers 2021*, 10, 143, doi: https://doi.org/10.3390/computers10110143.

[9] W. Liu, K. Lis, M. Salzmann and P. Fua, "Geometric and Physical Constraints for Drone-Based Head Plane Crowd Density Estimation," *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Macau, China, 2019, pp. 244-249, doi: 10.1109/IROS40897.2019.8967852.

[10] A. K. Pai, A. K. Karunakar and U. Raghavendra, "A Novel Crowd Density Estimation Technique using Local Binary Pattern and Gabor Features," *2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, Lecce, Italy, 2017, pp. 1-6, doi: 10.1109/AVSS.2017.8078556.

[11] Y. Zhu, K. Ni, X. Li, A. Zaman, X. Liu et al., "Artificial Intelligence Aided Crowd Analytics in Rail Transit Station," *Transportation Research Record*, p. 03611981231175156, 2023.

[12] L. Bai, C. Wu, F. Xie and Y. Wang, "Crowd density detection method based on crowd gathering mode and multi-column convolutional

neural network," Image and Vision Computing, vol. 105, p. 104084, 2021.

[13] A.A.A. Abalkhail and S.M.A. Al Amri, "Saudi Arabia's Management of the Hajj Season through Artificial Intelligence and Sustainability," Sustainability, vol. 14, no. 21, p. 14142, 2022.

[14] K. Rezaee *et al*., "IoMT-Assisted Medical Vehicle Routing Based on UAV-Borne Human Crowd Sensing and Deep Learning in Smart Cities," in *IEEE Internet of Things Journal*, vol. 10, no. 21, pp. 18529-18536, 1 Nov.1, 2023, doi: 10.1109/JIOT.2023.3284056.

[15] Q. Yu, L. Hu, B. Alzahrani, A. Baranawi, A. Alhindi and M. Chen, "Intelligent Visual-IoT-Enabled Real-Time 3D Visualization for Autonomous Crowd Management," in *IEEE Wireless Communications*, vol. 28, no. 4, pp. 34-41, August 2021, doi: 10.1109/MWC.021.2000497.

[16] G. Solmaz, P. Baranwal and F. Cirillo, "CountMeIn: Adaptive Crowd Estimation with Wi-Fi in Smart Cities," *2022 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, Pisa, Italy, 2022, pp. 187-196, doi: 10.1109/PerCom53586.2022.9762354.

[17] I. Ahmed, M. Ahmad, A. Ahmad, and G. Jeon, 'IoT-based crowd monitoring system: Using SSD with transfer learning', Computers & Electrical Engineering, vol. 93, p. 107226, Jul. 2021, doi: 10.1016/j.compeleceng.2021.107226.

[18] E. Prezioso, F. Giampaolo, S. Izzo, M. Savoia and F. Piccialli, "Integrating Object Detection and Advanced Analytics for Smart City Crowd Management," *2023 IEEE International Conference on Networking, Sensing and Control (ICNSC)*, Marseille, France, 2023, pp. 1-6, doi: 10.1109/ICNSC58704.2023.10318989.

[19] M. Al Duhayyim, E. Alabdulkreem, K. Tarmissi, M. Aliebreen and B.S.I.A. El khier *et al*., "Aquila Optimization with Transfer Learning Based Crowd Density Analysis for Sustainable Smart Cities", *Applied Sciences*, vol. 12, no. 21, p. 11187, Nov. 2022, doi: 10.3390/app122111187.

[20] J. R. Santana, L. Sánchez, P. Sotres, J. Lanza, T. Llorente and L. Muñoz, "A Privacy-Aware Crowd Management System for Smart Cities and Smart Buildings," in *IEEE Access*, vol. 8, pp. 135394-135405, 2020, doi: 10.1109/ACCESS.2020.3010609.

[21] D. A. Guastella, V. Campss and M. -P. Gleizes, "A Cooperative Multi-Agent System for Crowd Sensing Based Estimation in Smart Cities," in *IEEE Access*, vol. 8, pp. 183051-183070, 2020, doi: 10.1109/ACCESS.2020.3028967.

[22] Y. Yang, J. Yu, C. Wang and J. Wen, "Risk assessment of crowd-gathering in urban open public spaces supported by spatio-temporal big data," *Sustainability*, vol. 14, no. 10, p. 6175, 2022.

[23] B. Padmaja, B.J.D. Kalyani, V. Chapala, S. Solanki and A. Kaipa, "Crowd abnormal behaviour detection using convolutional neural network and bidirectional LSTM," *In AIP Conference Proceedings*, vol. 3007, 2024.

[24] W. Zhou, X. Yang, X. Dong, M. Fang, W. Yan and T. Luo, "MJPNet-S*: Multistyle Joint-Perception Network with Knowledge Distillation for Drone RGB-Thermal Crowd Density Estimation in Smart Cities," in *IEEE Internet of Things Journal*, doi: 10.1109/JIOT.2024.3369642.

[25] Z. Yang, D. Hu, Q. Guo, L. Zuo and W. Ji, "Visual E2C: AI-Driven Visual End-Edge-Cloud Architecture for 6G in Low-Carbon Smart Cities," in *IEEE Wireless Communications*, vol. 30, no. 3, pp. 204-210, June 2023, doi: 10.1109/MWC.019.2200518.

[26] S. Sahu, A.K. Singh, S.P. Ghrera and M. Elhoseny, "An approach for de-noising and contrast enhancement of retinal fundus image using CLAHE," Optics & Laser Technology, vol. 110, pp. 87-98, 2019.

[27] Y. Sun, J. Zhang, Z. Yu, Y. Zhang and Z. Liu, "The Bidirectional Gated Recurrent Unit Network Based on the Inception Module (Inception-BiGRU) Predicts the Missing Data by Well Logging Data," *ACS omega*, vol. 8, no. 30, pp. 27710-27724, 2023.

[28] M. Oszust, "Enhanced marine predators algorithm with local escaping operator for global optimization," *Knowledge-Based Systems*, vol. 232, p. 107467, 2021.

[29] S.M. Abdullah, M. Periyasamy, N.A. Kamaludeen, S.K. Towfek, R. Marappan et al., "Optimizing Traffic Flow in Smart Cities: Soft GRU-Based Recurrent Neural Networks for Enhanced Congestion Prediction Using Deep Learning," *Sustainability*, vol. 15, no. 7, p. 5949, 2023.

[30] K. Venkatrao and S. Kareemulla, "HDLNET: A Hybrid Deep Learning Network Model With Intelligent IOT for Detection and Classification of Chronic Kidney Disease," *IEEE Access*, vol. 11, pp. 99638-99652, 2023.

[31] L. Ma, F. Ma, W. Cao, B. Lou, X. Luo et al., "A Multi-Strategy Improved Sooty Tern Optimization Algorithm for Concrete Dam Parameter Inversion," *Water*, vol. 16, no. 1, p. 119, 2023.

[32] F. Alrowais, S.S. Alotaibi, F.N. Al-Wesabi, N. Negm, R. Alabdan et al., "Deep Transfer Learning Enabled Intelligent Object Detection for Crowd Density Analysis on Video Surveillance Systems," *Applied Sciences*, vol. 12, no. 13, p. 6665, 2022.