Semiconductor Optoelectronics
ISSN: 1001-5868

COATI-BELUGA WHALE HYBRID MODEL FOR HUMAN TRAJECTORY PREDICTION BASED ON SPATIO-TEMPORAL SEMANTIC FEATURES

N. Venkata SubbaReddy
Assistant Professor, Dept of CSE, SreeNidhi Institute of Science & Tech. Hyderabd-501 301, India. Email:nvsreddy@sreenidhi.edu.in.

Dr. D. S. R. Murthy
Professor, Dept of CSE, Anurag University, Hyderabad - 500088, India. Email: dsrmurthy.1406@gmail.com

Abstract

Humans have a natural perception of direction that takes, and predicting human trajectories accurately is crucial for ensuring the safety and efficiency of various systems, such as autonomous vehicles, robot navigation, and pedestrian flow management. However, present techniques for modeling this property for a range of applications, such as tracking pedestrian traffic and identifying unusual occurrences, are limited with the natural characteristics. Also, traditional methods involve significant computing costs for prediction process, particularly for long sequence predictions. In order to assist with more precise and effective trajectory predictions, this paper proposes a novel human trajectory prediction model with spatio-Temporal Semantic feature Extraction process. At first, the input video is processed for frame conversion and improve the image quality by using gaussian filter during preprocessing phase. Subsequently, some appropriate features are extracted from the preprocessed frames (consider as image). Features includes Spatial Binning feature, Histogram of Oriented Gradients feature, Improved LGXP feature and Spatial Temporal Semantic features are extracted in the feature extraction process. After the feature extraction process, prediction process takes place, where a hybrid model is proposed. The hybrid prediction model is the combination of Improved DCNN and SVM model. Also, to enhance the performance of the hybrid model, the weights of hybrid model is tuned by the proposed optimization algorithm namely, BMCOA (Beluga whale Merged Coati Optimization Algorithm). Finally, the simulation result demonstrates that the suggested technique outperforms other traditional techniques in terms of various metrices.

Keywords- Trajectory Prediction, Spatial-Temporal semantic features, Improved DCNN, Optimization, Deep Learning

1.0 Introduction

The automatic interpretation of human actions in video sequences has been a focus of artificial intelligence development. There are several beneficial applications, including autonomous driving, socially cooperating robotics, and sophisticated visual monitoring. However, despite the significant advancement's researchers have observed, it remains still

challenging to forecast human trajectories using automatic analysis [9] [10] [19].

Human behavioral deductive thinking, principles of common sense, cultural norms, and relationships with other people and the surroundings determine the human movement activities and trajectories [11] [12] [20]. People may accurately predict the instantaneous movements of others and react to them. However, It is quite difficult to predict the human movement by the automated system. To create an efficient human trajectory prediction method, the environment's physical limitations on the movement of humans must be modeled, along with other humans or automobiles' motions and their interpersonal behaviors [13] [14] [21] [22].

As an example, persons would prefer to move on the path than crossing a busy street. A person can change his trajectory by predicting the future movements of those surrounding them. When others do the same, their target gets affected. As people are so complicated, predicting human trajectory has become a very difficult problem. In the past few decades, human forecasting of trajectory has undergone significant advances due to the effective deep learning technique [15] [16] [23] [24].

In earlier approaches, understanding behavioral trends among moving participants like humans or automobiles and representing the linguistic features of the navigational surroundings are the main targets. To be able to forecast each agent's upcoming trajectory, recent methods concentrate on the relationships among all the people involved in the environment. RNNs based data-driven algorithms are currently used to simulate human-human conversations, evaluate societal acceptance, and predict the combined influence of any agent in the environment [17] [18] [25]. These problems therefore aim to create a significantly more effective learning and trajectory prediction system for people. Accordingly, this paper proposes a human trajectory prediction model, and the key contributions are summarized below:

- Proposing an Improved Local Gabor XOR Pattern feature extractor with new gradient computation with sobel operator along with Spatial Temporal semantic feature to define the occurrence at a specific time and location.
- Proposing a hybrid prediction model that combines IDCNN and SVM. In this prediction model IDCNN is proposed to improve the accuracy of the prediction model with the involvement of new layer, Improved Scalar Dot product Attention layer.
- Proposing a new optimization algorithm termed as BMCOA for tuning the optimal weights of hybrid model, which ensures better convergence than the conventional models.

The remaining section of this proposed work is described as follows: Literature review of the human trajectory prediction method is presented in Section 2.0. Proposed a new hybrid model with BMCOA is presented in Section 3.0. Simulation outcome of the proposed work is presented in Section 4.0. At last, the conclusion of this work is presented in Section 5.0.

2.0 Literature Review

In 2021, Abel Díaz Berenguer et al., [1] suggested a model which employs a latent variable in a framework that was cognizant about the relationship among people and environment for pedestrians' path prediction. In order capture human-contextual communication, the model suggested was dependent on relevant information which impacted walkers' paths. Throughout to decoding overall paths, researchers used an implicit variational

represent to clarify the ambiguity regarding their future paths and recorded comparative interpersonal effects between every person in the situation as well as their relationship with the scenario's set. In comprehensive tests utilizing publicly accessible datasets, it proved that the trajectory forecasting algorithm outperformed the latest algorithms through the use of environmental data as well as an unobserved variational framework.

In 2019, Jianqi Zhong et al., [2] pointed out that the way human movement happens within the 3D real world and the ways they behaved was well captured in the 3D space, it proved far easier to identify and forecast pedestrians paths in the 3D environment. In order to achieve this, DNNs used to identify and monitor the position of humans using a dual-camera setup. Such dual DNNs met their stereo uniformity condition throughout the estimation of poses. To forecast human motion trajectories via 2D to three-dimensional space, researchers modified the SocialGAN algorithm that was already in place. The experiments conducted showed that, in comparison with current latest techniques, the suggested strategy significantly enhanced pedestrians trajectory prediction accuracy.

In 2021, Dapeng Zhao and Jean Oh [3] created for learning, identify, and recover features in serial trajectory information, Social-PEC was the name assigned to the CNN-based method. Studies conducted on the human being's trajectory forecasting issue demonstrated that the algorithm executed was on level with the modern and advanced, as well as in certain instances even better. In addition, the suggested method provided a natural explanation for the procedure for making decisions, exposing the uncertainty in earlier use of a pooling layer.

In 2020, Zipei Fan et al., [4] suggested a one-time method of learning to address issues with changing users sets and inadequate information for human trajectory detection. A Siamese network was implemented to gather the criteria for recognition, and a slice encoding neural networks was developed to transmit the spatiotemporal attributes. Tests were carried out using an actual human motion database to demonstrate the algorithm's beneficial performance.

In 2023, Yue Yu et al., [5] presented a Bi-LSTM aided approach to predict the unpredictability of crowd-sourced pedestrian paths in complicated metropolitan contexts. Actual gait-length and direction variation variables within observed step interval were automatically included into the suggested movement uncertain modeling structure for human motion recognition. The attributes were then built as the starting point characteristics of the unpredictability forecasting algorithm by combining them with the position, quickly, and simulated direction data derived from the GNSS. Utilizing real-world data sets, comparisons with the state-of-the-art unpredictability modeling techniques were made. The findings showed an effective the superior performance of the provided Bi-LSTM helped structure to accomplish more precise and adaptive motion unpredictability forecasting, according to by several measurements. This work might encourage trustworthy analysis for crowdsourcing metropolitan large-scale data, as it offered a precise and useful method for modeling the motion unpredictability of human motions under complicated metropolitan surroundings.

In 2023, Mark Nicholas Finean et al., [6] proposeda innovative approach to human prediction of trajectory combines recognition of intentions with movement planning using trajectories optimisation. Utilizing real-time RGB-D sensor information collected by the onboard the camera, the resulting structure was tested on the Toyota HSR in outside conditions. They shared the Oxford-IHM information set, provided research on a publically accessible dataset, and showed cutting-edge results for predicting human motion. The Oxford-IHM

dataset was utilized to forecast human trajectories as people moved among areas of attraction inside a building. Although tracked by a motion-capture structure, the humans are being observed by stationary and robot-mounted RGB-D sensors.

In 2020, Xiaodong Zhao et al., [7] developed a system for attention to gather spatial affinities and to discover the value of past trajectories data at various time instants. A DNN was used to gauge the relative significance of the neighbours in order to investigate spatial affinities. The method demonstrated superior efficacy on publically accessible datasets, according to the experimental findings, when compared to modern technology.

In 2019, Chuanwei Ding et al., [8] suggested a unique DRDT approach based on FMCW radar method, to detect continuous movement of people under varied settings that simulate a real-life surroundings. Continuous movements might be broken up using this approach and considered as individual occurrences. The backscattered impulses were initially utilized to create range-Doppler structures, that were composed of several range-Doppler mappings. The DRDT was then taken using these images so as to track the movements of people in actual time throughout the duration, spectrum, and Doppler domain. Following that, every human movement on the DRDT image was identified and separated using the highest point searching technique. In order to create a ML classification algorithm, spectrum, Doppler radar, RCS, and dispersal characteristics have been extracted and integrated using a multi-domain fusing method. Even under difficult circumstances of the distance, viewing position, guidance, and human variety, this accomplished precise and consistent identification. Numerous tests have been run to demonstrate its viability and being superior, and the results for continuous categorization showed a median accuracy of 91.9%.

In 2023, Amin Manafi Soltan Ahmadi and Samaneh Hoseini Semnani [38] suggested a model for predicting human trajectory that includes a LSTM network. For that, To identify which elements of the information, the authors employ attention scores. Each input feature's attention score is determined, with a higher score reflecting that feature's greater significance in determining the outcome. These scores were initially computed for the location, speed and surrounding humans for the target human locations and motions of individuals. Using attentional ratings, this proposed model could order the most important data in the input information and provided more precise forecasts. Then extract the attention method generates attention scores, which incorporate to predict, enter them into the trajectory prediction module. After that, created a new neural layer that handles attention scores after extracting them in order to accomplish this adds positional data to before concatenating them. The authors assessed on the freely accessible ETH and UCY datasets and the FDE was a way to gauge a system's performance with metrics for ADE. Finally, demonstrate that the Comparison to the Social LSTM, the customized algorithm performs better in determining the course of pedestrians in congested areas spaces

In 2021, Mengyang Huang et al., [39] proposed a traffic network using a Bayonet-Corpus based on the context of traffic crossings. Additionally, the sequence of traffic intersections along a single trajectory is predicted using the Bi-GRU. Initially, this study presented a new traffic network modeling technique based on the context of traffic intersections because real traffic networks are typically complicated and disordered and cannot depict the higher dimensional link among traffic crossings. A Bayonet-Corpus was created from traffic intersections in real trajectory sequence, inspired by the probabilistic language model. This

approach blocks the complicated structure of the real traffic network and reconstructs the traffic network space by mapping the vehicle route nodes into a high-dimensional space vector. Next, a matrix was created by mapping the bayonets sequence in the real traffic network. We use Bi-GRU to bidirectionally model the trajectory matrix for the purpose of prediction since the trajectories sequence was bidirectional and it was handle input from both the forward and the backward directions simultaneously.

Table I: Features and challenges of existing techniques

Author [Citation]	Methodologies	Features	Challenges
Abel Díaz Berenguer et al., [1]	SSALVM	In this approach, the CNN-acquired scenario design details aid in modeling human-contextual interactions.	If it employed an ordered graph framework, it could be effective.
Jianqi Zhong et al., [2]	3D stereo human trajectory learning and prediction	It reduced the prediction error with a median of 47%, greatly increasing the trajectory's accuracy of predictions.	To address the overfitting problem, the suggested model needs to be tuned using an optimization approach.
Dapeng Zhao and Jean Oh [3]	Temporal CNN	By avoiding ambiguity in data collection, PEC was used, which improved the model's efficiency.	The prediction of human trajectory was not integrated into the physical constraints on the environment.
Zipei Fan et al., [4]	one-shot trajectory identification	In order to represent human trajectories, four distinct topologies of networks were investigated. The results revealed the slice CNN architecture exceeded the slice-Bi-LSTM architecture for more fuzzy classification.	It didn't employ a more complicated network architecture, and more complex negative sampling methods could not be employed to record more intricate trajectory interpretation.
Yue Yu et al., [5]	Bi-LSTM	By taking into account a series of	It needs optimization algorithm to adjust

		observed vectors rather than just immediately near data, the algorithm was able to change effectively.	the suggested model and address the overfitting issue
Mark Nicholas Finean et al., [6]	hybrid trajectory prediction	It might forecast and take into account human movement patterns in a robots's workspace using real-time sensor readings, allowing the robot to prevent collisions while executing whole-body movement planning in a range of difficult and dynamic conditions.	However, because the sensors used more power, the simulation was unable to perform real-time processing.
Xiaodong Zhao et al., [7]	ST-Attention model	It accurately forecasts the subsequent direction of the object of study and is completely differentiated. It also automatically recognizes the significance of previous historical trajectory at various time moments and weighs the impact of nearby individuals on the target's path.	It was unable to apply intelligent optimization techniques to acquire every parameter..
Chuanwei Ding et al., [8]	DRDT	In the suggested model, continuous movements of people might be located and precisely separated over a lengthy period of time with little	A real-time human motion detection system's viability would not be investigated.

		computation using a peak search strategy.	
Amin Manafi Soltan Ahmadi and Samaneh Hoseini Semnani [38]	LSTM	The proposed model improves by 6.2% and 6.3% in ADE and FDE in comparison to the Social LSTM outcomes	Need to involved advanced attention mechanism in the proposed model
Mengyang Huang et al., [39]	Bi-GRU	The suggested method obtained feaster in converges and more easy to obtained prediction	Have to considered various events for human trajectory prediction such as festivals, traffic accidents etc.

2.1 Problem Statement

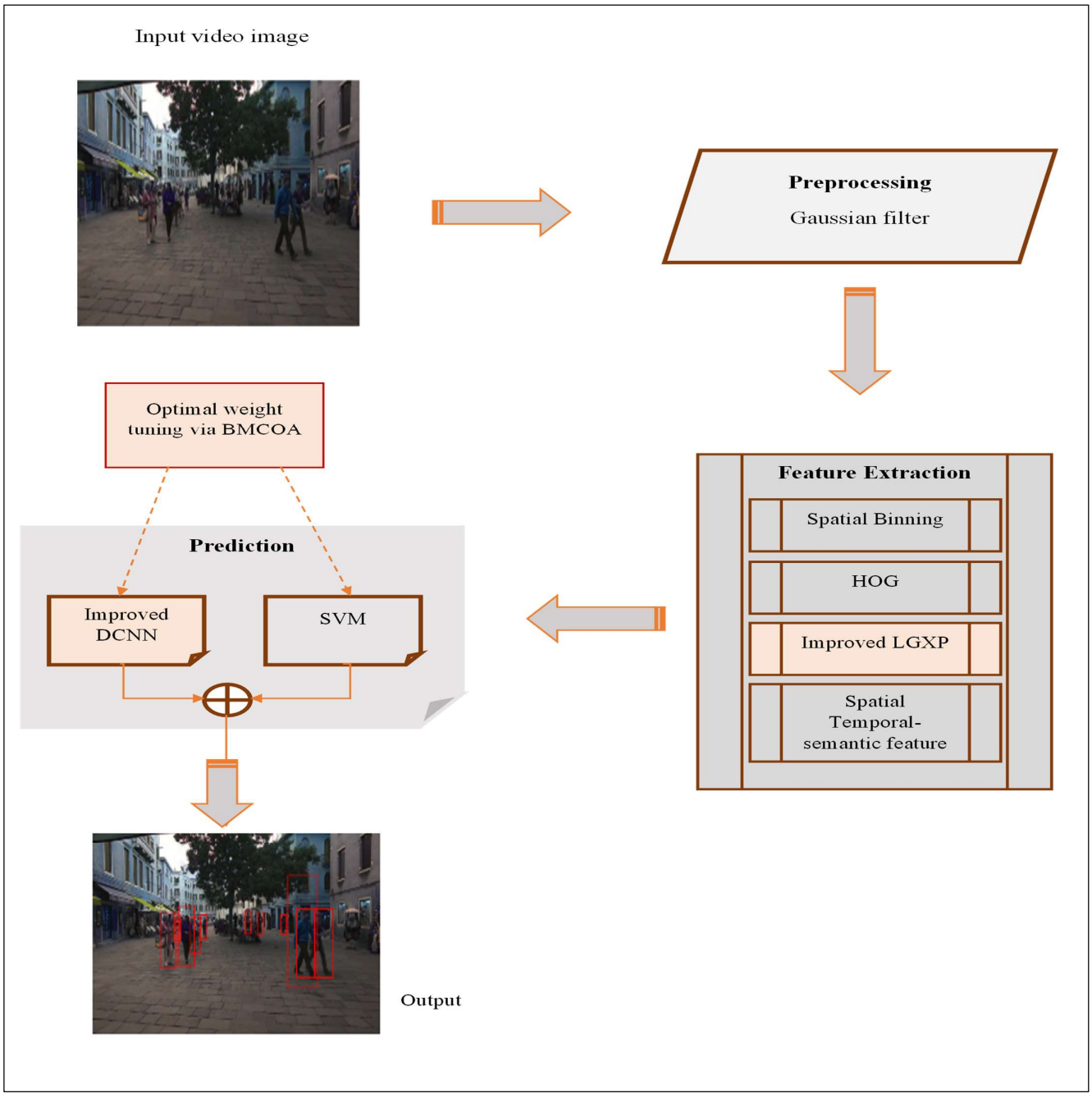
The challenge of human trajectory forecasting utilizing deep learning aims to predict people's planned pathways and positions in different situations. This crucial problem has implications in many domains, such as computer monitoring, self-driving cars, security for pedestrians, and interaction between humans and computers. The goal of the subject is to develop predictive algorithms which can evaluate past trajectory information—which is frequently obtained using detectors like camera or Satellite devices—and produce precise forecasts of people's expected future movements. This entails considering complex elements including the surrounding environment, interaction between pedestrians and objects, and unique patterns of behavior. Precise trajectory prediction is necessary to advance the development of cities, improve the safety and effectiveness of autonomous systems, and facilitate organic and simple human-computer interactions. Due to the inherent ambiguity and changing character of human mobility, it is still a difficult task. By identifying complex spatiotemporal correlations in trajectories information, deep learning approaches such as RNNs, CNNs and attention processes have showed promise in solving this issue. Some of the challenges in conventional methods are given in Table I.

3.0 Proposed Architecture for Human Trajectory Prediction

Accurate modeling and forecasting behavior among people still remain a challenging part due to its complexities in areas including irregular occurrence identification, movement calculation, and behavior prediction. This paper proposes an optimally trained hybrid model for human trajectory prediction with Spatio Temporal Semantic feature extraction process. Three steps are included in this paper for human trajectory prediction model, which are explained as follows.

- Preprocessing is the initial step for prediction model. Here, the input video is converted into frames (image) and enhances the quality of the image by suing gaussian filter process

- Then the feature extraction process is carried out from the preprocessed image. Here, the features like Spatial Binning features, HOG, Improved LGXP, Spatial Temporal Semantic features are extracted in this feature extraction stage.
- Finally, the prediction step is takes place by a hybrid model that combines IDCNN and SVM. In this hybrid model, the optimal weight is tuning through BMCOA (Beluga Whale Merged Coati Optimization Algorithm) to get the higher accuracy of the prediction results. Figure 1 shows the overall architecture of the prediction model.



estimated or determined by using Eq.(1). Where, D denotes the normalization constant, σ denotes the standard deviation of the gaussian kernel. Therefore, the preprocessed image is denoted as I_F^P

$$GF(x,y)=\frac{1}{D}e^{\frac{x^2+y^2}{2\sigma^2}} \tag{1}$$

3.2 Feature Extraction $\overline{F}_T$

After the preprocessing phase, feature extraction process is takes place from the preprocessed image. According to the proposed work, Spatial Binning feature, Histogram of Oriented Gradients feature, Improved LGXP and Spatial Temporal-semantic feature are extracted and the final feature set is defined as $\overline{F}_T$.

3.2.1 Spatial Binning $\overline{F}_{SB}$

Using OpenCV, this paper implemented spatial binning [27] to the preprocessed image. The size of the preprocessed image was reduced using OpenCV's "cv2.resize()" function, and the 1D vector of features was created using numpy's "ravel()" function. Analyzing the regularity of quantitative information categorized into groups covering a range of potential values is the goal of binning. The extracted spatial binning feature is represented as $\overline{F}_{SB}$.

3.2.2 Histogram of Oriented Gradients (HOG) $\overline{F}_{HOG}$

HOG is a feature identifier for identifying objects in artificial intelligence and processing images. This characteristic identifier is a description of an image's contents or an image patches that strips away unneeded data to create a simpler image. The method records the number of times a change in direction appears in specific areas of an image.

According to this paper, the preprocessed image's upper parts are identified in the desired regions obtained by background elimination with HOG [28] since upper bodies are definitely seen even though bottom bodies are concealed by additional items. The backdrop image is then modified based on the results of the detection and the gathering of background statistics. P represent the intensity of the pixels of the surrounding area in this method and it is expressed in Eq.(2). Where, P' represents the average time of the pixel intensity, ω represents the frequency , σ represents the amplitude of intensity, t represents the time, C represents the coefficient $(-1 \le C \le 1)$, ξ represents the highest value of the noise. Three processes is carried out in each pixel depending on the results of the detection. The first process is considered as if P is not a part of any rectangles that are being detected and satisfied the condition as $P'-\sigma-\xi \le P \le P'+\sigma+\xi$. At this case, it is determined that the pixel is present in the background, and P' and $\bar{\sigma}$ are modified depends on Eq.(3) and Eq.(4). Here, N represents the parameter of the modified speed. The second process is considered as if P is not in the rectangular region, then it is not satisfied the condition of $P'-\sigma-\xi \le P \le P'+\sigma+\xi$. At this situation, it is determined that the pixel is located in the area of a moving objects other than a human, and instead of updating P', $\bar{\sigma}$ is modifies as per Eq.(5) wherein m represents the

parameter determining the updated speed in the object areas ($m \geq n$). Thus, moving objects besides people can eventually be incorporated into the background. The image that serves as the background is modified regularly to avoid the detector from producing unexpected errors. Finally, the third process is in the same position, P is shown in rectangles. The pixel is recognizing as existing in human regions in this instance and also, P' and σ are not updated in these areas. By maintaining these quantities, the identified areas remain the subtraction areas and do not appear in the background image. The HOG features extraction is denoted as $\overline{F}_{HOG}$.

$$P = P' + \sigma \sin\left(2\pi\omega t\right) + C\xi \tag{2}$$

$$\overline{P}' = \frac{\left(n-1\right)P' + P}{N} \tag{3}$$

$$\overline{\sigma} = \frac{\left(n-1\right)\sigma + \sqrt{2\left(P-P'\right)^2}}{N} \tag{4}$$

$$\overline{\sigma} = \frac{\left(m-1\right)\sigma + \sqrt{2\left(P-P'\right)^2}}{m} \tag{5}$$

3.2.3 Improved LGXP $\overline{F}_{ILGXP}$

In this paper, an improved LGXP [29] feature descriptor is proposed. The proposed concept includes five main stages, such as Image Acquisition, Noise Suppression, Improved gradient computation, LGXOR and Histogram of image. Figure 2 depicts the flowchart of improved ILGXP process.

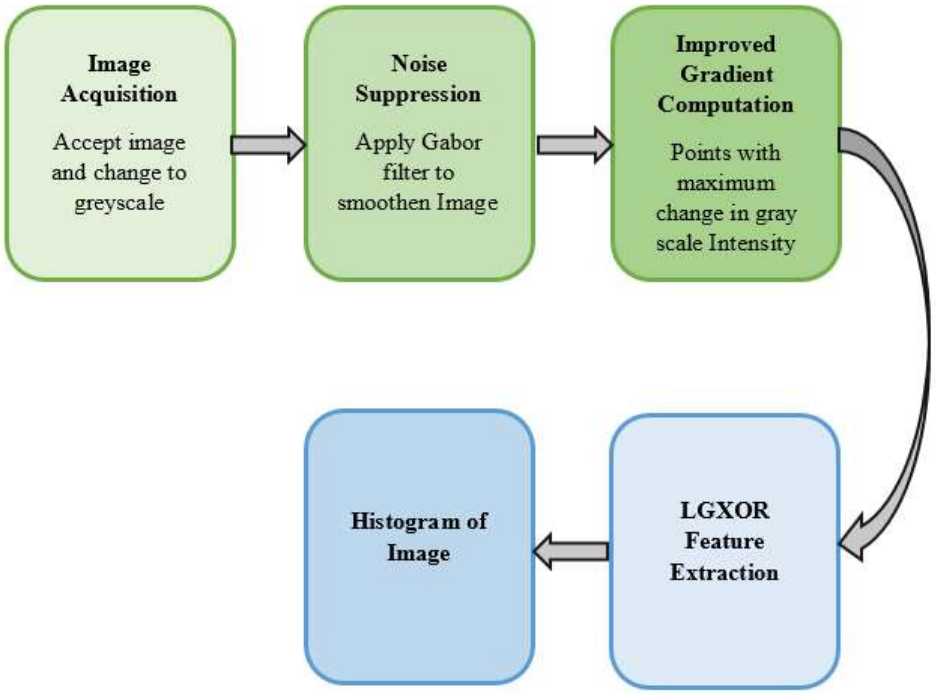


Figure 2: Flowchart of improved ILGXP process

(i) Image Acquisition

The input image (preprocessed image, I_F^P) is changed to a non-colored image referred to as a grayscale image for simple analysis. The image's intensity can be found between 0 and 255. The mathematical representation of grayscale image is explained as follows:

Pixels arranged in rows and columns defines the image. The resolution of an image defines the width of each column and row. The total number of pixels in an image, denoted as nx, ny represents the dimension of the image. An intensity matrix of pixels at the coordinates (xi, yi) of the digital image can utilized for representing it, here nx indicating the width and ny indicating the height. Let's consider that grayscale image, $P_{ij}\left(i=1,2,3,..,n,\ j=1,2,3,..,n\right)$ where P_{ij} indicates the intensity of the digital image for a position (i, j). The image has fluctuating noise N_{ij} that affects the pixel at location (i, j). As a result, a combined setting, as indicated in Eq.(6), can be used to describe the intensity of the image's gray level. The gray scale image is represented as $G(x, y)$

$$G_{i,j} = P_{i,j} + N_{i,j} \tag{6}$$

(ii) Noise Suppression via Gabor filtering

Initially the gray scale image $G(x, y)$ is taken as input into the Gabor filter. The application of this combination has shown positive outcomes. Gabor filters are band pass filters that let through the specified range of frequencies. Gabor filters are particularly sensitive to the intersection of two differing intensity in an image within a specified place and direction.

A 1-D Gabor filter is able to be visualized mathematically as the combination of a sinusoidal function stated by Eq.(7) in a Gaussian function. Where f and φ are the frequency and phase offsets of the sinusoidal wave, respectively, and g and σ_x are the x-axis Gabor function and standard deviation in that direction.

$$g(x) = \exp\left(-\frac{x^2}{2\sigma_x^2}\right)\exp\left(i\left(\frac{2\pi x}{f}+\varphi\right)\right) \tag{7}$$

$$e^{ix} = \cos x + i\sin x,\ \left(i=\sqrt{-1}\right)$$

Since the image represents a 2-D function, Eq.(7) has been modified to Eq.(8). Here, the real and imaginary parts are represented as $g(x, y) = \exp\left(-\frac{x^2 + s^2 y^{i^2}}{2\sigma^2}\right)\cos\left(\frac{2\pi x^i}{f}+\varphi\right)$ and

$g(x, y) = \exp\left(-\frac{x^2 + s^2 y^{i^2}}{2\sigma^2}\right)\sin\left(\frac{2\pi x^i}{f}+\varphi\right)$,respectively.where,

$x^i = x\cos\theta + y\sin\theta;\ y^i = -x\sin\theta + y\cos\theta$.

$$g(x, y) = \exp\left(-\frac{x^2 + s^2 y^{i^2}}{2\sigma_x^2}\right)\exp\left(i\left(\frac{2\pi x^i}{f}+\varphi\right)\right). \tag{8}$$

The grayscale image is mathematically modeled with the Gabor filter combination to filter eliminate undesirable data. σ describes the filter would blur the image. Assuming that

the filter's input and output images are $G(x, y)$ and $G'(x, y)$, respectively, Figure 3 shows a possible smoothing layout according to Eq.(8).

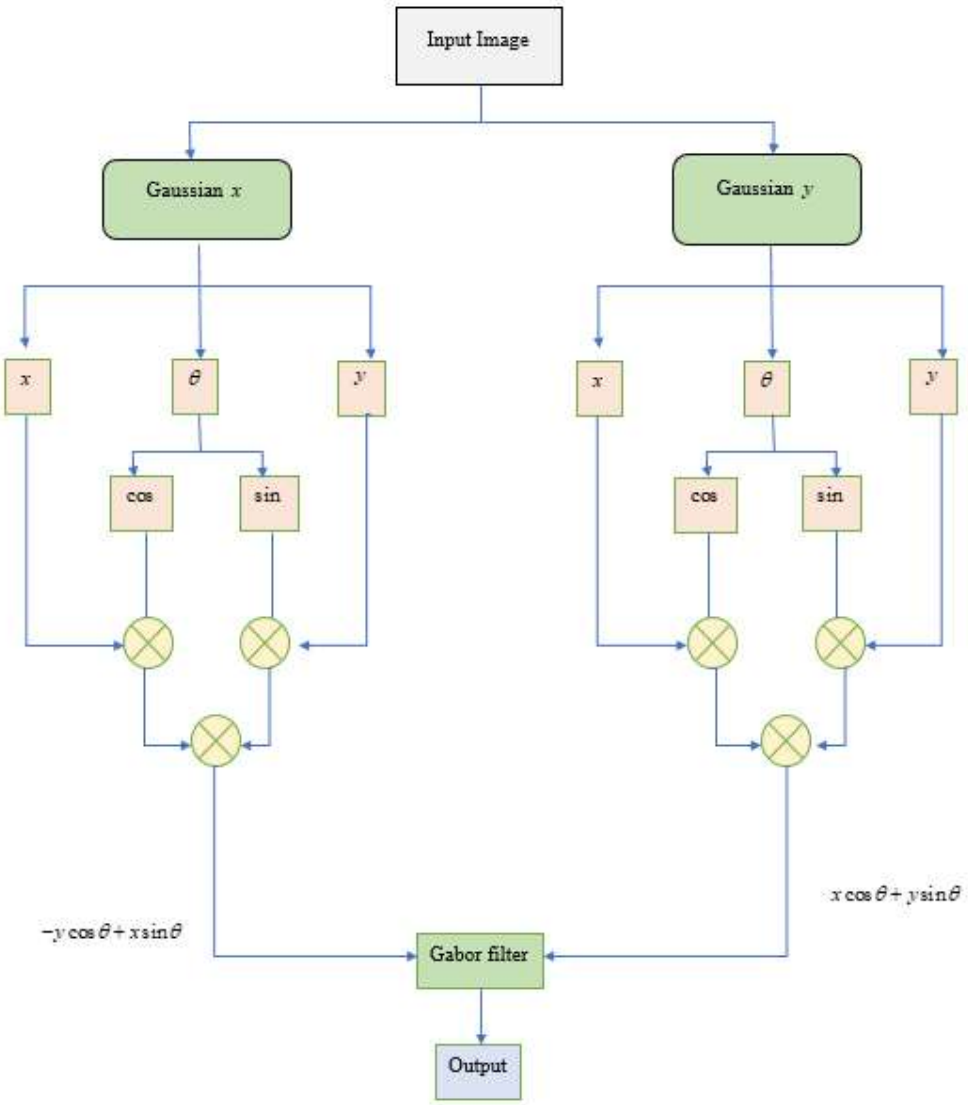


Figure 3: Layout of Noise Smoothing based on Gabor filter

(iii) Improved gradient Computation

Calculating the gradient and identifying the pixel spots with the greatest change in grayscale intensity are the primary issues. To overcome this, in this work, the method known as the Sobel operator is used to calculate the pixel gradient in the image. The operator is used in the i and j directions of 2-D space. These two directions' sobel operators are provided by the matrices M_i and M_j.

The orientations j of the matrix M_i and M_j are convolved with blurred image pixels

$G'(i,j)^i$. Here, M_i and M_j, each values are calculated by using formula, which is given in Eq.(9) and Eq.(10), respectively, T,B and L,R denotes the top, bottom and left right of the image pixels.

$$M_i = \left[\frac{x_i - \left(x_i^T + x_i^B\right)/2}{\left(x_i - x_i^L\right) + \left(x_i - x_i^R\right)} - \frac{1/2\left(x_c - x_i^T\right) + \left(x_c - x_i^B\right)}{\left(x_c - x_i^L\right) + \left(x_c - x_i^R\right)} \right] \tag{9}$$

$$M_j = \left[\frac{x_j - \left(x_j^T + x_j^B\right)/2}{\left(x_j - x_j^L\right) + \left(x_j - x_j^R\right)} - \frac{1/2\left(x_c - x_j^T\right) + \left(x_c - x_j^B\right)}{\left(x_c - x_j^L\right) + \left(x_c - x_j^R\right)} \right] \tag{10}$$

Next the gradients at pixel point (i,j) is calculated as per Eq.(11) .

$$\begin{aligned} g_i &= M_i * G'(i,j) \\ g_j &= M_j * G'(i,j) \end{aligned} \tag{11}$$

The magnitude of Edge E is computed by Eq.(12), where n represents the total number of data points

$$E = \sqrt{\left(\frac{\sum_{i=1}^{n} \left| g_i^{\,2} - g_j^{\,2} \right|}{n} \right)} \tag{12}$$

(iv) LGXOR feature

After the calculation of improved gradient, LGXOR [30] features are extracted. To identify the tracks, their Gabor individuality is quite efficient and adequate. The band pass as well as Gabor filters have been employed for the feature extraction. As all the characteristic features had not been accessible from the identical orientation, a set of filters is successfully used with separating orientation to obtain frequency information and subsequently the characteristics at distinct orientations. Scaling is done to get the most frequent data for each direction. Stated differently, the extraction of the highest frequency data is considered in both the scale and the orientation. The stages in the LGXP are initially categorized into diverging varieties, and then the middle pixel's quantization stages constituted in which the LXP operator worked, and each of its neighbors. Finally, the local representation is created by concatenating the significant binary variables combined within the main image pixel.

(v) Histogram of Image

The last stage of the Improved LGXP is obtaining the histogram of image[31]. Here the histogram of image is obtained from the output of the LGXOR extracted features. These features are divided into a l number of non-overlapping sub blocks with all scales (ν) and orientation (φ) of the image, which is represented in the Eq.(13). here, $H_{\nu,\varphi,j}$ represents the histogram of the j^{th} sub block of LGXOR map $(j = 1,2,..,l)$ .Consider two LGXOR descriptors for human trajectory H^1 and H^2 are evaluated as per Eq.(14). Eq.(15) indicates the histogram intersection operation, here L represents the total number of histogram bins. Thus, the features of improved LGXOR are finally denoted as $\overline{F}_{ILGXP}$.

$$H=\left[H_{\varphi 0,\nu 0,1},...,H_{\varphi 0,\nu 0,l};H_{\varphi 0-1,\nu 0-1,l},..,H_{\varphi 0-1,\nu 0-1,l}\right] \tag{13}$$

$$S'\left(H^1,H^2\right)=\sum_{\varphi=\varphi 0}^{\varphi_{\varphi-1}}\sum_{\nu=\nu 0}^{\nu_{s-1}}\sum_{j=1}^{l}\cap\left(H^1_{\nu,\varphi,j},H^2_{\nu,\varphi,j}\right) \tag{14}$$

$$\cap\left(H^1,H^2\right)=\sum_{j=1}^{L}\min\left(h_i^1,h_i^2\right) \tag{15}$$

3.2.4 Spatial Temporal-Semantic feature $\overline{F}_{STS}$

The spatial Temporal feature is represented as $\overline{F}_{STS}$. Initially, analyze a file containing video with OpenCV in order to get data regarding the video frames. Then, compute the location-based attributes for every image as repeatedly cycle across every image. After that, calculate the HOG features in this instance. The list 'feature vectors' contains the feature matrices for each frame. Then, arrange the spatial feature vectors to create "spatial features" after processing each frame. To create a single spatial-temporal representation of features, additional temporal grouping or evaluation on the "spatial features" object are evaluated. Then, determine the spatial-temporal feature shapes.

Thus, the overall feature extraction from the preprocessed image is denoted as $\overline{F}_T$, $\overline{F}_T=\left[\overline{F}_{SB}\ \overline{F}_{HOG}\ \overline{F}_{ILGXP}\ \overline{F}_{STS}\right]$

3.3 Prediction of Human Trajectory via BMCOA trained Hybrid Model

According to this paper, prediction is the final stage, a hybrid model with BMCOA optimization is proposed for the human trajectory prediction. The hybrid model is the combination of Improved DCNN and SVM model. The optimal weight of the hybrid model is tuned by the BMCOA method which combines Beluga Whale and Coati Optimization Algorithm. The process of hybrid model is as follows: The extracted features $\overline{F}_T$ are subjected to both the Improved DCNN and SVM model simultaneously, and the obtaining prediction scores are averaged together to obtain the final prediction score of the human trajectory prediction model. This section explained the human trajectory prediction process.

3.3.1 Improved DCNN

An improved DCNN [32] is proposed in this prediction hybrid model. Initially, the overall feature $\overline{F}_T$ is taken as input into the network. In the improved DCNN model, the weight W_1 is optimally tuned by using BMCOA algorithm. Deep learning can be identified from other types of NNs simply by taking into account its hidden layer element. The current CNNs typically consist of just one input layer, one output layer, including one hidden layer. However, DL uses a number of hidden layers to produce decisions that are more precise. It is necessary to determine the extent to which hidden, or the total number of layers that data flows through when going through various identification algorithms. Based on the preliminary output layer, each layer of the network is trained using different features in this way. One example of feed forward is CNN. A collection of neurons with biases and weights is used to build ANNs and

CNNs. Each neuron processes information and performs the dot product. Single vectors are fed into neural networks, which then process them through a number of hidden layers.

Neurones make up each individual placed layer, and they are each fully coupled to all of the others in the layer before it. The output layer, which is the last fully connected layer, has responsibility for outsourcing the prediction scores as the ultimate decision. The input feature $\overline{F}_T$ is passed to the convolutional layer 1. After each convolutional layer 1(filter size 32 and kernel size 1), apply ReLU Activation function to reduce gradient problem in the system. Then the feature map is forwarded to batch Normalization layer to generate the map.

Then the output from the convolution layer 1 is passed to the convolution layer 2 (second block) (filter size 64 and kernel size 1) to obtain feature set. Here again apply the ReLU function to reduce the gradient problem. After that, the feature map is forwarded to batch Normalization to generate the map.

Next, it is passed to the Improved Scalar Dot product Attention layer as shown in Eq.(16). The mapping of the query (input) and a collection of key-value combinations to an output, wherein the query, keys, values, and output all consist of vectors, is known as an attention function. The result is calculated as the weighted total of the values, with each value's weight determined by the compatibility relationship between the input values with a key-corresponding search. The main advantages of the attention layer are the long input sequences can be handled by attention techniques [33] because they enable the model to choose concentrate on the most important portions of the input. In addition, by giving a visual representation for the portions of the input the model is focusing on, they can be utilized to make a model easier to understand.

$$Attention\left(Q,K,V\right)=Soft\max\left(\sigma\left(\frac{Q_k^T\times Q_k}{\left(\sqrt{d_k}+Q_k\right)/d_k}\right)v\right) \tag{16}$$

Where, Q_k^T represents the two matrices, Q and K, σ represents the standard deviation, d_k represents the dimension of the key, v represents the actual values behind the keys.

Then the outcome from the attention layer is passed to the min-max pooling layer and further it passed to the flatten layer and finally it sends to the fully connected layer. Then the classification outcome by passing a vector to fully connected layer. Consequently, a dropout layer of 0.5 is added after the fully connected layer to avoid overfitting. Then the next layer is the improved activation function of Com-H-Sin(softmax) layer, which is expressed in Eq. (17). It is simpler to understand the outputs of the NN. The neural network's unprocessed output is converted into a vector of probabilities via the Com-H-Sin(softmax) activation function [34], which is simply the distribution of probabilities across the categorization. Where β=0.1. Then the output layer of the Improved DCNN gives the prediction score. Figure 4 shows the architecture of Improved DCNN.

$$Com-H-\mathrm{Sin}\,e(soft\max)=\left[\left(\mathrm{Sinh}\left(\beta x\right)+\mathrm{Sinh}^{-1}\left(\beta x\right)\right)*\left[\frac{\left(\exp(x_i)\right)}{\left(\frac{\exp\left(x_i\right)}{\sum_j \exp\left(x_j\right)}\right)}\right]\right] \quad (17)$$

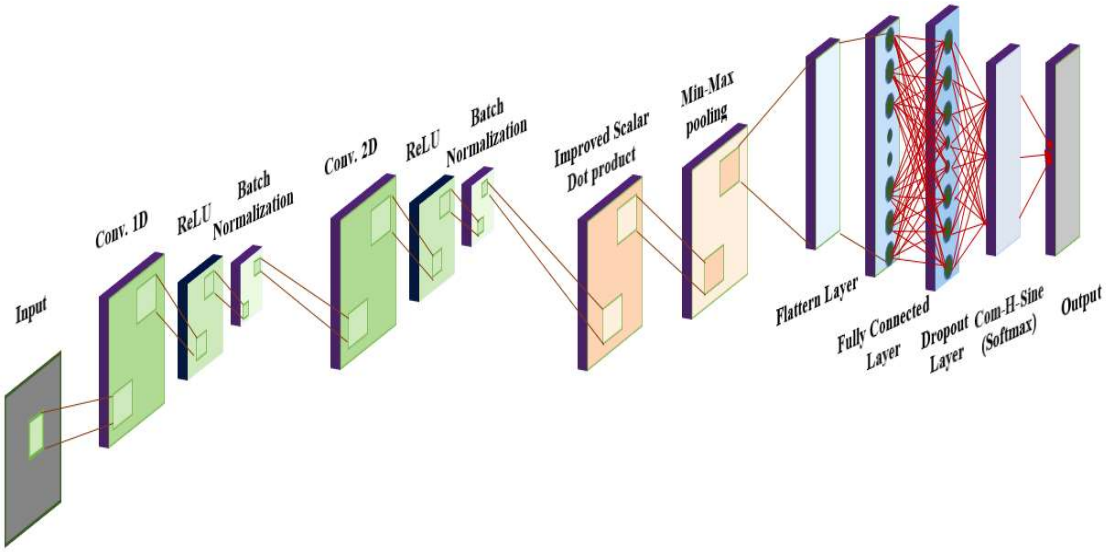


Figure 4: Architecture of Improved DCNN

3.3.2 SVM

With a few modest modifications, SVM [35] is a flexible method that may be used for classification and regression issues. SVM depends on identical concepts as in categorization for regression-based tasks, with a few minor variations. Here, the input data is taken as $\overline{F}_T$. W_2 is the weight of the SVM classifier, which is optimally tuned by using BMCOA algorithm. A kernel function is used to translate the input information to a high-dimensional, nonlinear space, thereby making it easier to identify a hyperplane in the data. For prediction purposes, a variety of kernel-based functions, such polynomials, RBF, the sigmoid and linear functions, among others, can be used. Here used the RBF kernel method to predict for the purposes of this investigation. The parameter for regularization of the SVM architecture was configured. The aforementioned choices were made in order to precisely maximize the SVM model's performance. The prediction score is obtained from the SVM model. Fig 5 shows the architecture of SVM.

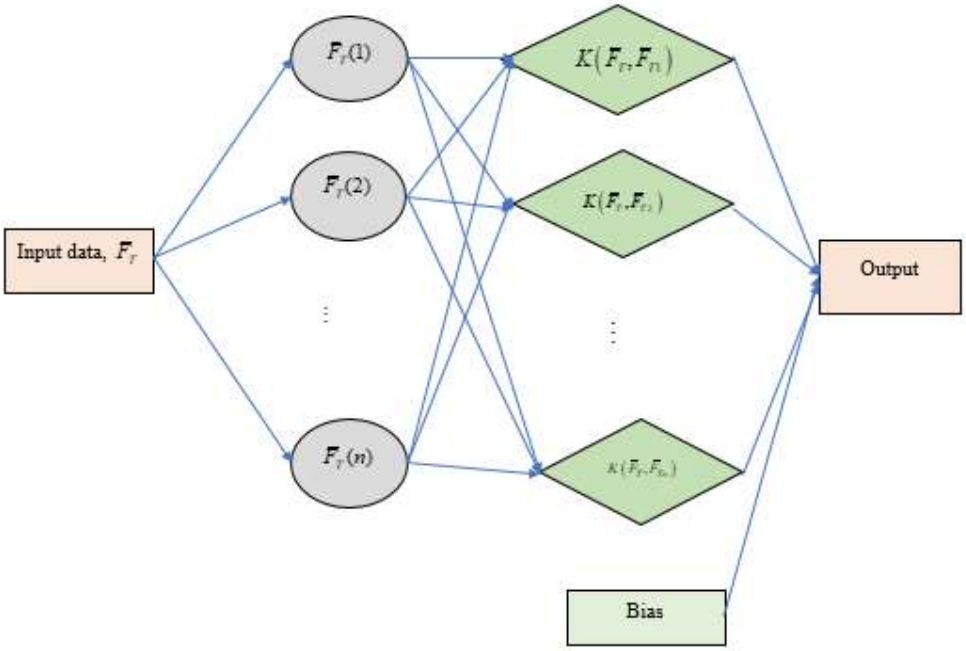


Figure 5: Architecture of SVM

Finally, outcome from both the classifiers are combined to averaged together to obtain the final prediction score of the human trajectory.

3.3.3 Tuning of optimal weight via proposed BMCOA algorithm

3.3.3.1 Solution encoding

The solution given to the BMCOA are weights of the hybrid model, denoted as $W_i = [W_1 \;\; W_2]$. Here, W_1 represents the weight of the IDCNN and W_2 represents the weight of SVM. Here, the population size of the proposed work is 25, and also fix the upper bound and lower bound value is 1 and 0 respectively.

3.3.3.2 Objective function

$$\hat{F} = \min\left(MSE\right) \tag{18}$$

The objective function of this proposed model is indicated as $\hat{F}$. It is defined as the minimum of MSE, which is expressed as per Eq.(18). MSE considers the error between actual value and predicted value.

3.3.3.3 Mathematical model of BMCOA

According to this paper, the mathematical model of the BMCOA is explained as follows. In this algorithm, the coati exploration phase [36] is updated by using the exploitation phase in the Beluga whale optimization [37]. Here the position and their attitudes of coati is considered as the weight of the hybrid model. Initially, the location of the weight in the search space is set at randomized at the start of the BMCOA execution as per Eq.(19). Where, $i = 1,2,..,T$, $j = 1,2,...,n$, W_i represents the i th position of the weight in the search space, $w_{i,j}$ represents the j th decision variable values. T represents the total number of weight to be considered in the algorithm, n represents the total number of decision variables, R denotes the real number at the interval 0 to 1. U_j^B represents the upper bound of the j th decision

variable and L_j^B represents the lower bound of the j th decision variable. Eq.(20) shows the mathematical representation of weight in population matrix. The positioning of the potential solutions in choice variables results in the estimation of various values for the problem's objective function. Eq. (21) is used to display these values, which is already mentioned in Eq.(18). Where $\hat{F}$ denotes the objective function value that is obtained as per i th position of the weight.

$$W_i : w_{i,j} = L_j^B + R \cdot \left(U_j^B - L_j^B \right) \tag{19}$$

$$W = \begin{bmatrix} W_1 \\ \vdots \\ W_i \\ \vdots \\ X_T \end{bmatrix}_{T \times n} = \begin{bmatrix} w_{1,1} & \cdots & w_{1,j} & \cdots & w_{1,n} \\ \vdots & \ddots & \vdots & \iddots & \vdots \\ w_{i,1} & \cdots & w_{i,j} & \cdots & w_{i,n} \\ \vdots & \ddots & \vdots & \iddots & \vdots \\ w_{T,1} & \cdots & w_{T,j} & \cdots & w_{T,n} \end{bmatrix}_{T \times n} \tag{20}$$

$$\hat{F} = \begin{bmatrix} \hat{F}_1 \\ \vdots \\ \hat{F}_i \\ \vdots \\ \hat{F}_T \end{bmatrix}_{T \times 1} = \begin{bmatrix} \hat{F}(W_1) \\ \vdots \\ \hat{F}(W_i) \\ \vdots \\ \hat{F}(W_T) \end{bmatrix}_{T \times 1} \tag{21}$$

The value of the objective function serves as the benchmark for candidate solution quality in metaheuristic algorithms. The population member who results in the evaluation of the best value for the objective function is referred to as the best population weight. The best weight of the population is updated in each iteration of the algorithm since the weight solutions are updated at that time. It consists of two stages in BMCOA, such as Updating exploration stage and exploitation stage.

Stage 1: Proposed Exploration stage

In this stage, the iguana hunting and attack tactics are considered for the position of optimal weight is chosen. It is fully based on the coati hunting and attacking behavior, Here, the best position of weight is assumed to the iguana's position. Also, it's believed that half of the coatis ascend the tree while the other half waits for the iguana to fall to the ground. This position the optimal weight for this stage is calculated as per Eq.(22). Where, $i = 1, 2, .., \left\lfloor \frac{T}{2} \right\rfloor$, $j = 1, 2, .., n$.

$$W_i^{P1} : w_{i,j}^{P1} = w_{i,j} + R \cdot \left(G_j - S \cdot w_{i,j} \right) \tag{22}$$

$$w_{i,j}^{P1} = w_{i,j} + R \cdot G_j - R \cdot S \cdot w_{i,j} \tag{23}$$

At this stage, the exploitation phase of the beluga whale is merged in the Eq.(23). In the beluga whale optimization, (BWO), the levy flight strategy is introduced and the mathematical expression of the position based on the beluga catches the prey with the levy flight strategy in Eq.(24). Then rearrange the equation and obtained the current positon of

weight as per Eq.(25).Where, K represents the current iteration, W_i^K ,W_R^K , W_{best}^K represents the current position of the i th weight, random weight and optimal (best) weight, R_1, R_2 represents the random number between 0 and 1, Q_1 represents the random jump strength. I_{LF} represents the levy flight function and it calculated as per Eq.(26), σ is defined as per Eq.(27), c and d represents the distributed random function.

$$W_i^{K+1} = R_1 W_{best}^K - R_2 W_i^K + Q_1 \cdot I_{LF} \cdot \left(W_R^K - W_i^K \right) \tag{24}$$

$$W_i^{K+1} = R_1 W_{best}^K - R_2 W_i^K + Q_1 \cdot I_{LF} \cdot W_R^K - Q_1 \cdot I_{LF} \cdot W_i^K$$

$$W_i^{K+1} = R_1 W_{best}^K - W_i^K \left(R_2 + Q_1 \cdot I_{LF} \right) + Q_1 \cdot I_{LF} \cdot W_R^K$$

$$R_1 W_{best}^K - W_i^K \left(R_2 + Q_1 \cdot I_{LF} \right) + Q_1 \cdot I_{LF} \cdot W_R^K = W_i^{K+1} \tag{25}$$

$$-W_i^K \left(R_2 + Q_1 \cdot I_{LF} \right) = W_i^{K+1} - R_1 W_{best}^K - Q_1 \cdot I_{LF} \cdot W_R^K$$

$$W_i^K = -\left[\frac{W_i^{K+1} - R_1 W_{best}^K - Q_1 \cdot I_{LF} \cdot W_R^K}{\left(R_2 + Q_1 \cdot I_{LF} \right)} \right]$$

$$I_{LF} = 0.05 \times \frac{c \times \sigma}{\left| d \right|^{1/\beta}} \tag{26}$$

$$\sigma = \left(\frac{\Gamma\left(1+\beta\right) \times \sin\left(\pi\beta/2\right)}{\Gamma\left(\left(1+\beta\right)/2 \right) \times \beta \times 2^{(\beta-1)/2}} \right)^{1/\beta} \tag{27}$$

Substitute W_i^K value in Eq.(23) using Eq.(28) and get the updated position of BMCOA as per Eq.(29)

$$w_{i,j}^{P1} = -\left[\frac{W_i^{K+1} - R_1 W_{best}^K - Q_1 \cdot I_{LF} \cdot W_R^K}{\left(R_2 + Q_1 \cdot I_{LF} \right)} \right] + R \cdot G_j + R \cdot S \cdot \left[\frac{W_i^{K+1} - R_1 W_{best}^K - Q_1 \cdot I_{LF} \cdot W_R^K}{\left(R_2 + Q_1 \cdot I_{LF} \right)} \right] \tag{28}$$

$$w_{i,j}^{P1} = -\frac{W_i^{K+1}}{R_2 + Q_1 \cdot I_{LF}} + \frac{R_1 W_{best}^K}{R_2 + Q_1 \cdot I_{LF}} + \frac{Q_1 \cdot I_{LF} \cdot W_R^K}{R_2 + Q_1 \cdot I_{LF}} + R \cdot G_j + \frac{R \cdot S \cdot W_i^{K+1}}{R_2 + Q_1 \cdot I_{LF}} - \frac{R \cdot S \cdot R_1 W_{best}^K}{R_2 + Q_1 \cdot I_{LF}} - \frac{R \cdot S \cdot Q_1 \cdot I_{LF} \cdot W_R^K}{R_2 + Q_1 \cdot I_{LF}}$$

$$w_{i,j}^{P1} + \frac{W_i^{K+1}}{R_2 + Q_1 \cdot I_{LF}} - \frac{R \cdot S \cdot W_i^{K+1}}{R_2 + Q_1 \cdot I_{LF}} = \frac{R_1 W_{best}^K}{R_2 + Q_1 \cdot I_{LF}} + \frac{Q_1 \cdot I_{LF} \cdot W_R^K}{R_2 + Q_1 \cdot I_{LF}} + R \cdot G_j - \frac{R \cdot S \cdot R_1 W_{best}^K}{R_2 + Q_1 \cdot I_{LF}} - \frac{R \cdot S \cdot Q_1 \cdot I_{LF} \cdot W_R^K}{R_2 + Q_1 \cdot I_{LF}}$$

$$w_{i,j}^{P1}\left[1 + \frac{1}{R_2 + Q_1 \cdot I_{LF}} - \frac{R \cdot S}{R_2 + Q_1 \cdot I_{LF}}\right] = \frac{1}{R_2 + Q_1 \cdot I_{LF}}\begin{bmatrix} R_1 W_{best}^K + Q_1 \cdot I_{LF} \cdot W_R^K - R \cdot S \cdot R_1 W_{best}^K \\ -R \cdot S \cdot Q_1 \cdot I_{LF} \cdot W_R^K + \\ R \cdot G_j \left(R_2 + Q_1 \cdot I_{LF}\right) \end{bmatrix}$$

$$w_{i,j}^{P1} = \frac{\frac{1}{R_2 + Q_1 \cdot I_{LF}}\begin{bmatrix} R_1 W_{best}^K + Q_1 \cdot I_{LF} \cdot W_R^K - R \cdot S \cdot R_1 W_{best}^K \\ -R \cdot S \cdot Q_1 \cdot I_{LF} \cdot W_R^K + R \cdot G_j \left(R_2 + Q_1 \cdot I_{LF}\right) \end{bmatrix}}{\left[1 + \frac{1}{R_2 + Q_1 \cdot I_{LF}} - \frac{R \cdot S}{R_2 + Q_1 \cdot I_{LF}}\right]}$$

(29)

In this position, the random position of weight around the search area is calculated. As per Eq.(30) and Eq.(31). If the updated position is increasing the value of the objective function, the weight remains its previous iteration; if not, the new position is calculated. Eq. (32) is used to simulate for $i = \left\lfloor \frac{T}{2} \right\rfloor + 1, \left\lfloor \frac{T}{2} \right\rfloor + 2, ..., T$, $j = 1, 2, .., n$, and this update condition is for those values. Here, $\hat{F}_i^{P1}$ denotes its objective function value, R denotes a random real number in the interval [0, 1], G denotes the iguana's position in the search space, which essentially refers to the position of the optimal weight, and W_i^{P1} denotes the new position calculated for the ith iteration. Its jth dimension is called G_j, S represents an integer that is chosen at random from the set $\{1, 2\}$, G_j^{grd} denotes the randomly generated position of the iguana on the ground, $\hat{F}_{G^{grd}}$ denotes its objective function value.

$$G^{grd} : G_j^{grd} = L_j^B + R \cdot \left(U_j^B - L_j^B\right),\ j = 1, 2, ..., n$$

(30)

$$W_i^{P1} : w_{i,j}^{P1} = \begin{cases} w_{i,j} + R \cdot \left(G_j^{grd} - S \cdot w_{i,j}\right), & \hat{F}_{G^{grd}} < \hat{F}_i, \\ w_{i,j} + R \cdot \left(w_{i,j} - G_j^{grd}\right), & else \end{cases}$$

(31)

$$W_i = \begin{cases} W_i^{P1}, & \hat{F}_i^{P1} < \hat{F}_i, \\ W_i, & else \end{cases}$$

(32)

Stage 2: Exploitation stage

In this stage, the process of running away from predators are considered. Here, a coati flees from its location when it is attacked by a predator. Because of Coati's actions in this strategy, it is at a secure location near its current location, demonstrating the BMCOA's capacity for local search exploitation.

In order to replicate this behavior, a random position is created using Eq. (33) and Eq.(34). If the newly computed position increases the value of the objective function, which this condition uses to simulates, then it is acceptable as per Eq.(35). In this case, W_i^{P2} represents the new position based on the second phase of BMCOA, $w_{i,j}^{P2}$ denotes its jth dimension, $\hat{F}_i^{P2}$ denotes its objective function value; R denotes a random number in the interval [0, 1]; t denotes the iteration counter; LL_j^B and LU_j^B represents the local lower bound and local upper bound of the j th decision variable, respectively; L_j^B and U_j^B represents the lower and upper bound of the j th decision variable, accordingly.

Thus, BMCOA's iteration is finished once all weight in the search space in their positions changed depending on the first and second stages. Up until the algorithm's final iteration, the update of the population based on Eqs. (29) to (33), is repeated. When the BMCOA run is over, the output is the best optimal weight that could be obtained after all algorithm iterations. Figure 6 shows the flowchart of the BMCOA.

$$LL_j^B = \frac{L_j^B}{t}, LU_j^B = \frac{U_j^B}{t}, \quad t = 1,2,...,K \tag{33}$$

$$W_i^{P2}: w_{i,j}^{P2} = w_{i,j} + (1-2R)\cdot\left(LL_j^B + R\cdot\left(LU_j^B - LL_j^B\right)\right), \quad i = 1,2,..,V, \quad j = 1,2,...,n \tag{34}$$

$$W_i = \begin{cases} W_i^{P2}, \hat{F}_i^{P2} < \hat{F}_i, \\ W_i, \quad else \end{cases} \tag{35}$$

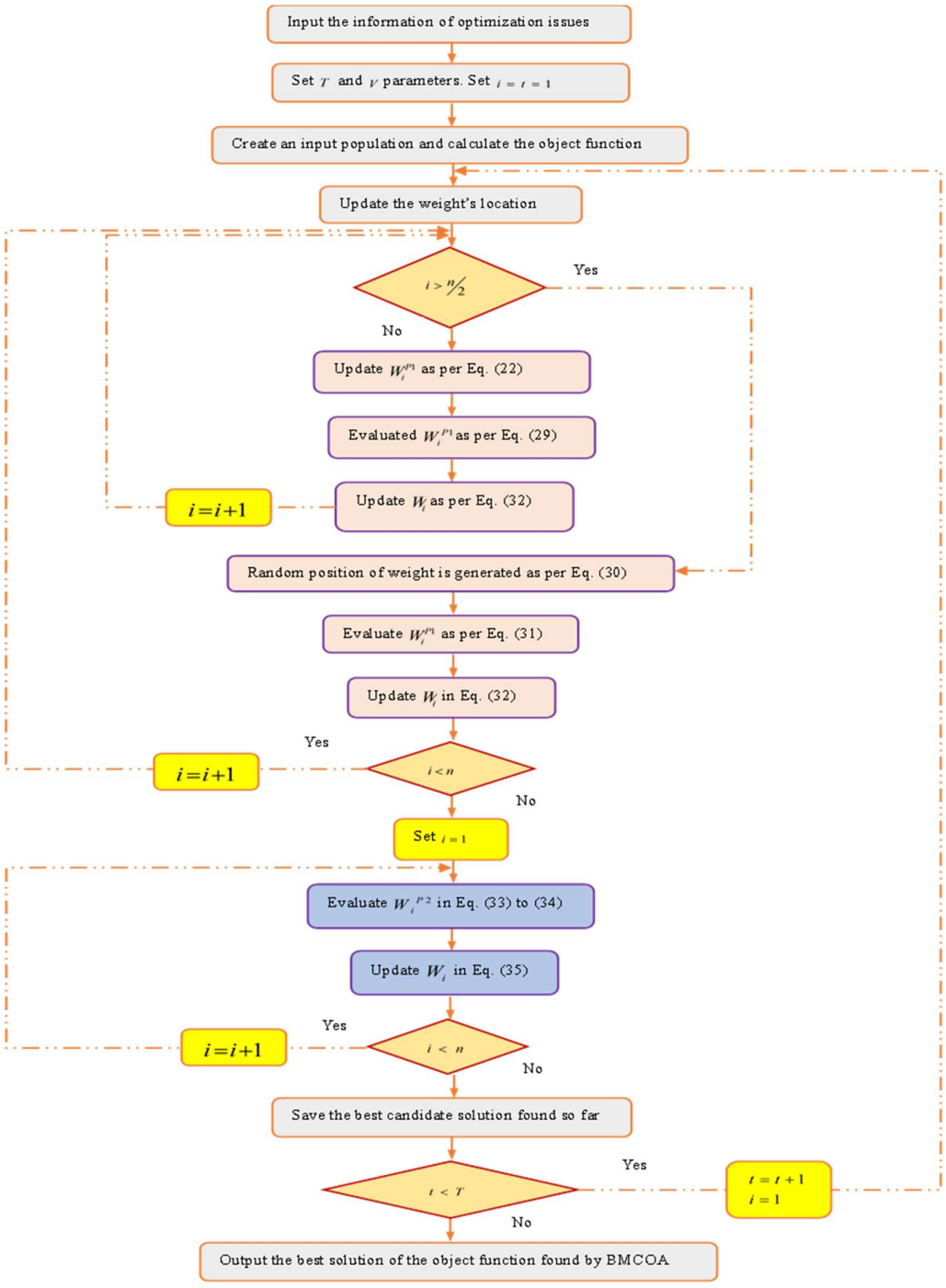


Figure 6: Flowchart of BMCOA

4.0 Results and Discussion

4.1 Experimental Setup

The Hybrid model+BMCOA model is implemented by PYTHON 3.7.9 in intel core i5 processor with 16GB RAM and the end-result was successfully verified. The Hybrid model+BMCOAmodel is compared with conventional approaches such as Coati Optimization Algorithm, Archimedes optimization algorithm (COA), Salp Swarm Algorithm (SSA), Hunger

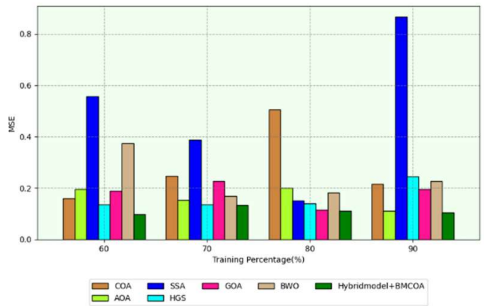
Games Search (HGS), Grasshopper Optimization Algorithm (GOA), Beluga Whale Optimization Algorithm (BWO), LSTM [38] and Bi-GRU [39]. Dataset used for training and evaluation is MOT17Det [40]

4.2 Dataset Description

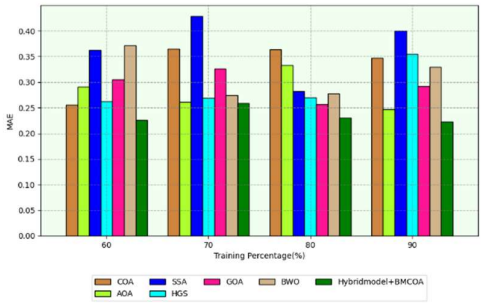
MOT17Det is extracted from the Multiple Object Tracking Benchmark. This framework was developed for the objective assessment of multiple people monitoring algorithms. Dataset contains a sizable number of datasets, some of which are already in use, and some brand-new difficult sequences. All the sequences have been detected already in the dataset. It also contains evaluation tool that offers a variety of metrics, including recall, precision, and running time. A simple approach to assess how well modern tracking techniques operate was provided in the dataset. Dataset contains training set and testing set. All the video sequence is in 1920x1080 resolution.

4.3 Error Analysis

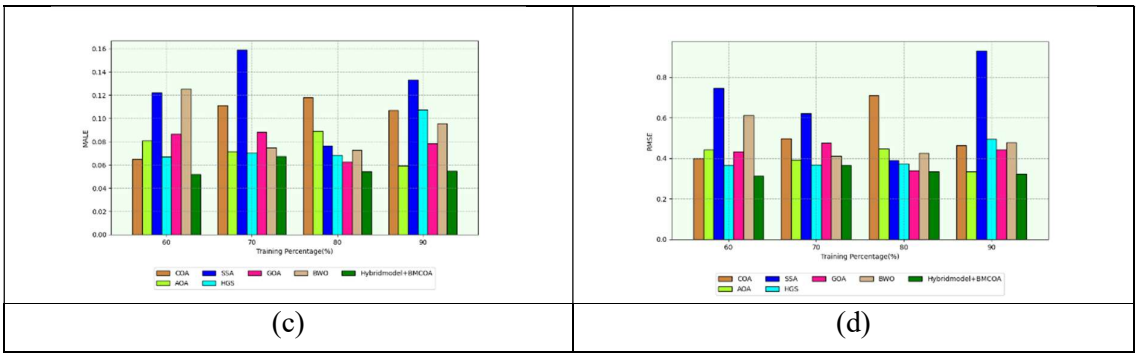
The Hybrid model+BMCOA model for Human Trajectory Prediction is analyzed based on error occurrence while prediction. If the error occurred decreased, the prediction of human trajectory will be more accurate. MSE, MAE, MALE, RMSE are the error metrices used to analyze the Hybrid model+BMCOA model for Human Trajectory Prediction shown in Fig 7. To analyze the Hybrid model+BMCOA model about the error occurrence, some conventional optimization algorithm like COA, AOA, SSA, HGS, GOA, BWO are compared to the Hybrid model+BMCOA model. The Hybrid model+BMCOA optimization has lesser error than other conventional techniques, MSE value is 0.105 for the Hybrid model+BMCOA model at learning percentage 90, which shows that the predicted value is similar to the exact value. For MAE, conventional optimization algorithm like COA algorithm has 0.346 error, AOA algorithm has 0.246 error, SSA algorithm has 0.400 error, hgs algorithm has 0.354 error, goa algorithm has 0.291 error BWO algorithm has 0.329 error at learning percentage 90, whereas Hybrid model+BMCOA optimization algorithm has less error of 0.222 error as shown in fig 7(b). Hybrid model+BMCOA with optimization algorithm obtains 0.067 as MALE value and RMSE value is 0.365 at learning percentage 70, which are comparatively lesser than the conventional algorithms. As the Hybrid model+BMCOA model converts the video input to enhanced frame image and Improved Local Gabor XOR Pattern (LGXP), Spatial Temporal- semantic feature are extracted from the pre-processed frame. Hybrid model+BMCOA model consists of IDCNN and SVM and applying hybrid optimization algorithm consists of COA and Blue Whale optimization algorithm helps the Hybrid model+BMCOA model to predict the human trajectory with less error.



(a)



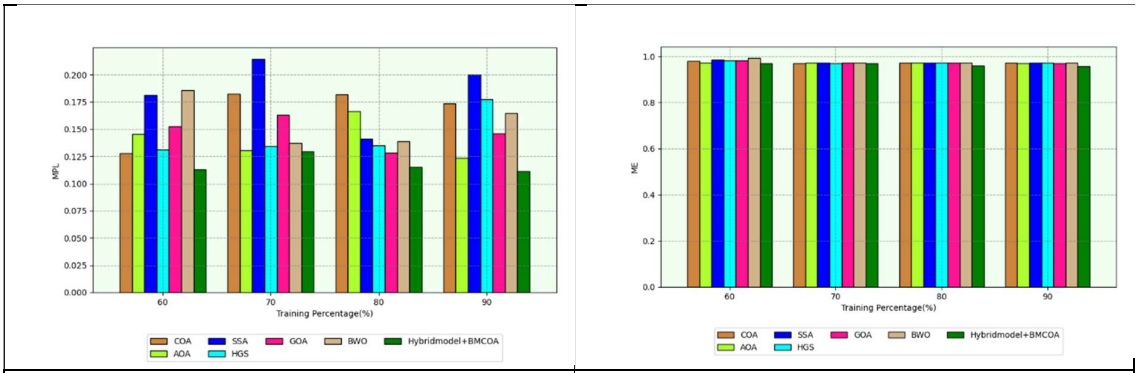
(b)



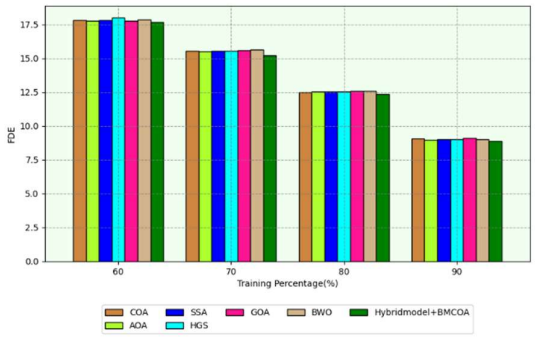
(c) (d)

Figure 7: Error Detection analysis of the developed optimization algorithm in comparison to extant schemes for (a) MSE (b) MAE (c) MALE (d) RMSE

Figure 8 represents the graphical representation of error detection analysis of the developed optimization algorithm in comparison to extant schemes for MPL, ME, FDE measures. As from the fig 8, all the error detection metrices shows that the Hybrid model+BMCOA model attains less error in different learning percentage, 60, 70, 80, 90. In the Hybrid model+BMCOA model, spatial temporal feature is extracted in the feature extraction phase, which helps the model to extract low-level information from image attributes in order to form a transfer and mapping of low-level information to high-level semantics. This makes the model error less so the prediction of human trajectory will be higher compared to other conventional approaches. In fig 8(a) MPL error value is 0.111 at learning percentage 90, whereas other techniques have 2% higher error than the Hybrid model+BMCOA model. FDE scheme is approximately same for conventional and proposed, as the Hybrid model+BMCOA model has 15.21 FDE value at learning percentage 70. As the learning percentage increases from 60 to 90, the error occurrence of the Hybrid model+BMCOA model with hybrid optimization algorithm becomes lesser.



(a) (b)



(c)

Figure 8: Error Detection analysis of the developed optimization algorithm in comparison to extant schemes for (a) MPL (b) ME (c) FDE

4.4 Comparative Analysis

To analyze the prediction performance of the Hybrid model+BMCOA model, conventional methods are compared to the proposed model. SVM, RF, KNN, DEEPMAXOUT, LSTM and Bi-GRU are the conventional techniques compare with the Hybrid model+BMCOA model. Comparison is done based on some error metrices like MSE, MAE, MALE, RMSE, MPL, ME, FDE to evaluate the prediction of Hybrid model+BMCOA model with conventional approaches. Table II represents the comparative analysis of Hybrid model+BMCOA approach to previous methods. From the table II it is clearly shown that the proposed hybrid model has less error than the pre-existing approaches. MSE value of the LSTM is 0.293, SVM is 0.229, RF is 0.112, Bi-GRU is 0.192, KNN is 0.229, DEEPMAXOUT is 0.122 and finally, the Hybrid model+BMCOA model has 0.105 as MSE value, which is lesser than all other conventional technique. Hybrid model+BMCOA model has 0.111 for MPL, where all other pre-existing approach has more error than proposed model and FDE value is 8.87, all other methods has more than 10 error value. ME is 0.956 occurred for the Hybrid model+BMCOA model, where other pre-existing model has higher error than the Hybrid model+BMCOA model, specifically DEEPMAXOUT technique has 2.18 ME error. MALE value obtained for the Hybrid model+BMCOA model is 0.054, which is very less compared to other conventional technique. Hybrid model+BMCOA model has less error than the conventional technique to predict the human trajectory. Error occurrence is lesser in Hybrid model+BMCOA model as it combines IDCNN and SVM and applyes hybrid optimization algorithm.

Table II Comparative Analysis of Hybrid model+BMCOA approach to previous methods

	LSTM [38]	SVM	RF	Bi-GRU [39]	KNN	DEEPMAXOUT	Hybrid Model+BMCOA
MSE	0.293411	0.229218	0.112683	0.192577	0.22914	0.122898	0.105044
MAE	0.290489	0.310245	0.393904	0.31245	0.283124	0.280303	0.22287
MALE	0.05904	0.097179	0.08371	0.078605	0.98475	0.085412	0.054681
RM	0.35563	0.3709	0.4012	0.354266	0.3442	0.328777	0.324105

SE	2	31	5		77		
MPL	0.145245	0.264682	0.146952	0.144519	0.141562	0.194015	0.111435
ME	1.522921	1.538569	1.878589	1.51114	1.774017	2.186717	0.956885
FDE	19.7736	11.06871	10.59825	19.68497	9.336915	23.59001	8.871124

4.5 Performance Analysis

In this section, the obtained image results are shown in figure 9. Also, the performance of proposed work is evaluated in terms of ablation study by varying model without feature extraction, model without optimization, model with conventional LGXP. The obtained results are given below.

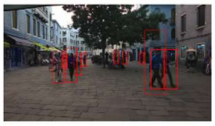
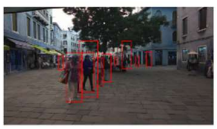
Original	Detected Image
	
	
	
	

Figure 9: Image results (a) original image, (b) Detected image

4.5.1 Ablation Study

Ablation study of the Hybrid model+BMCOA model is shown in the table III. Model without feature extraction, model without optimization, model with conventional LGXP are compared to the proposed approach. The error occurrence of the Hybrid model+BMCOA model is evaluated by MSE, MAE, MALE, RMSE, MPL, ME, FDE. MSE occurrence is 0.105 for the Hybrid model+BMCOA model where model without feature extraction, model without optimization and model with conventional LGXP has error higher than 0.3. Model without feature extraction obtains 0.077 MALE value, where the Hybrid model+BMCOA model has 0.054 MALE value, which shows that the Hybrid model+BMCOA model has higher performance than the model without feature extraction. In the Hybrid model+BMCOA model improved Local Gabor XOR Pattern (LGXP), Spatial Temporal- semantic feature are extracted, it makes the Hybrid model+BMCOA model to decrease the error in prediction of human trajectory. The model without optimization and model with conventional LGXP are similar in RMSE in the range of 0.38, where the Hybrid model+BMCOA model has less RMSE of 0.324. Hybrid model+BMCOA model has 8.87 FDE value, where other approach has 10% more error than the Hybrid model+BMCOA model. From the ablation study, it is proved that the Hybrid model+BMCOA model is highly performed for human trajectory prediction than the conventional techniques and the model without optimization and without feature extraction.

Table III: Ablation study of Hybrid model+BMCOA approach

	Model withoutfeature extraction	Model without optimization	Model with conventional LGXP	Hybrid model+BMCOA
MSE	0.327587	0.344469	0.336028	0.105044
MAE	0.31024	0.381363	0.345802	0.22287
MALE	0.077122	0.08853	0.082826	0.054681
RMSE	0.401	0.385599	0.393299	0.324105
MPL	0.26	0.361816	0.310908	0.111435
ME	0.998753	1.88905	1.443902	0.956885
FDE	17.98531	18.04094	18.01312	8.871124

4.6 Statistical Analysis

The statistical analysis of the Hybrid model+BMCOA model and conventional technique is conducted based on the occurrence of errors. The algorithm is executed multiple times to accurately determine the performance results. According to Table IV, the error occurrence of the Hybrid model+BMCOA model ranges from 0.097 to 0.133. In comparison, the proposed hybrid optimization algorithm achieves significantly lower error rates than the conventional technique. The minimum error occurrence of the Hybrid model+BMCOA model is 0.097, while the maximum error occurrence is 0.133. The mean value occurred by the Hybrid model+BMCOA model is 0.112, which is lower than that of the conventional techniques, where all other has mean above 0.16. Additionally, the standard deviation of the Hybrid model+BMCOA model is 0.013, whereas other techniques have values above 0.1. It is observed that models with lower error rates exhibit higher performance in terms of prediction accuracy. This observation is evident from the performance results of the Hybrid model+BMCOA work compared to the conventional technique.

Table IV: Statistical Analysis of Hybrid model+BMCOA approach to previous methods

	COA	AOA	SSA	HGS	GOA	BWO	Hybrid model+BMCOA
Mean	0.281888	0.165237	0.490495	0.164184	0.182225	0.23872	0.112016
Median	0.231175	0.174385	0.472285	0.13823	0.192713	0.205216	0.108431
Std	0.132639	0.035579	0.26009	0.046922	0.040792	0.081521	0.013323
Min	0.159986	0.112162	0.151706	0.134884	0.116164	0.169677	0.097772
Max	0.505216	0.200013	0.865705	0.245392	0.22731	0.374772	0.133427

4.7 Convergence Analysis

Convergence analysis is done to evaluate the cost value of the BMCOA model compared to conventional approaches. Proposed BMCOA model is iterated by changing the iteration count as 0, 5, 10, 15, 20, 25 as shown in figure 10. The BMCOA model has very low-cost value than the conventional technique such as COA, AOA, SSA, HGS, GOA, BWO results in higher performance of human trajectory prediction. The cost-value of the BMCOA model is indicated by pink line in the graph shown below. The cost-value of the BMCOA model initially starts from 0.0224 and in 5^{th} iteration the cost-value decreased to 0.0222. cost value is stable till the 18^{th} iteration and suddenly falls to approximately 0.0210 at the final stage of iteration. BMCOA model has the lowest of all other previous techniques, which clearly states that the BMCOA model has higher performance than the conventional techniques with lest convergence almost in all iterations.

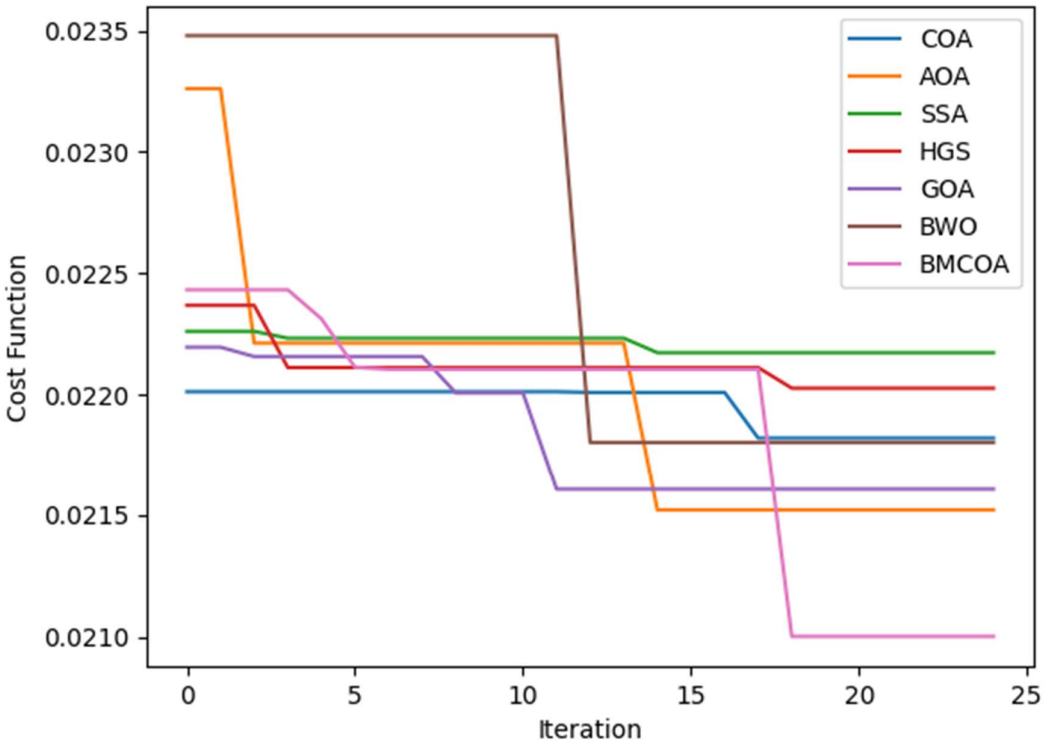


Figure 10: Convergence analysis of provided hybrid approach to previous methods

5.0 Conclusion

In this paper, a novel spatio-temporal semantic feature extraction process-based human trajectory prediction model was suggested. The input video is initially turned into a frame, and during the preprocessing stage, a gaussian filter is used to enhance the image quality. The preprocessed data was then used to extract a few relevant features. During the extraction procedure, features such as spatial temporal semantic features, improved LGXP feature, Histogram of Oriented Gradients feature, and spatial binning feature were extracted. Following feature extraction, a prediction procedure was conducted. Here, a hybrid model was suggested to improve the human trajectory prediction model's superior performance. The improved DCNN and SVM models were combined to create the hybrid prediction model. BMCOA techniques, or beluga whale Merged Coati Optimization Algorithm, was used to optimize both models' optimal weights. As a result, to acquire the final prediction score, the prediction score from each hybrid prediction model was obtained and averaged. This method was implemented using Python. The final results have been compared with numerous other algorithms. The mean absolute error, mean square log error, maximum error, mean pinball loss, distance, and mean square error were among the performance measures that were employed. The minimum and maximum error occurrence of the Hybrid model+BMCOA model was 0.097 and 0.133, respectively. The simulation result shows that the proposed method performs better on different metrics compared to other conventional methods.

REFERENCE

[1] A. Díaz Berenguer, M. Alioscha-Perez, M. C. Oveneke and H. Sahli, "Context-Aware Human Trajectories Prediction via Latent Variational Model," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 31, no. 5, pp. 1876-1889, May 2021, doi: 10.1109/TCSVT.2020.3014869.

[2] J. Zhong, H. Sun, W. Cao and Z. He, "Pedestrian Motion Trajectory Prediction With Stereo-Based 3D Deep Pose Estimation and Trajectory Learning," in IEEE Access, vol. 8, pp. 23480-23486, 2020, doi: 10.1109/ACCESS.2020.2969994.

[3] D. Zhao and J. Oh, "Noticing Motion Patterns: A Temporal CNN With a Novel Convolution Operator for Human Trajectory Prediction," in IEEE Robotics and Automation Letters, vol. 6, no. 2, pp. 628-634, April 2021, doi: 10.1109/LRA.2020.3047771.

[4] Fan, Z., Song, X., Chen, Q. et al. Trajectory fingerprint: one-shot human trajectory identification using Siamese network. CCF Trans. Pervasive Comp. Interact. 2, 113–125 (2020). https://doi.org/10.1007/s42486-020-00034-2

[5] Yue Yu, Yepeng Yao, Zhewei Liu, Zhenlin An, Biyu Chen, Liang Chen, Ruizhi Chen, "A Bi-LSTM approach for modelling movement uncertainty of crowdsourced human trajectories under complex urban environments". International Journal of Applied Earth Observation and Geoinformation, Vol. 122, August 2023, 103412, https://doi.org/10.1016/j.jag.2023.103412

[6] Mark Nicholas Finean, Luka Petrović, Wolfgang Merkt, Ivan Marković, Ioannis Havoutis, "Motion planning in dynamic environments using context-aware human trajectory prediction". Robotics and Autonomous Systems, Vol. 166, August 2023, 104450, https://doi.org/10.1016/j.robot.2023.104450

[7] X. Zhao, Y. Chen, J. Guo and D. Zhao, "A spatial-temporal attention model for human trajectory prediction," in IEEE/CAA Journal of Automatica Sinica, vol. 7, no. 4, pp. 965-974, July 2020, doi: 10.1109/JAS.2020.1003228.

[8] C. Ding et al., "Continuous Human Motion Recognition With a Dynamic Range-Doppler Trajectory Method Based on FMCW Radar," in IEEE Transactions on Geoscience and Remote Sensing, vol. 57, no. 9, pp. 6821-6831, Sept. 2019, doi: 10.1109/TGRS.2019.2908758.

[9] L. Wang, N. H. C. Yung and L. Xu, "Multiple-Human Tracking by Iterative Data Association and Detection Update," in IEEE Transactions on Intelligent Transportation Systems, vol. 15, no. 5, pp. 1886-1899, Oct. 2014, doi: 10.1109/TITS.2014.2303196.

[10] M. Antal, K. Buza and N. Fejer, "SapiAgent: A Bot Based on Deep Learning to Generate Human-Like Mouse Trajectories," in IEEE Access, vol. 9, pp. 124396-124408, 2021, doi: 10.1109/ACCESS.2021.3111098.

[11] Doi Thi Lan and Seokhoon Yoon, "Trajectory Clustering-Based Anomaly Detection in Indoor Human Movement". Sensors, Vol. 23, issue 6, 3318, 2023, https://doi.org/10.3390/s23063318

[12] Ali Abbasi, Sandro Queirós, Nuno M. C. da Costa, Jaime C. Fonseca and João Borges, "Sensor Fusion Approach for Multiple Human Motion Detection for Indoor Surveillance Use-Case". Sensors, Vol. 23, 3993, 2023, https://doi.org/10.3390/s23083993

[13] Tharindu Fernando, Simon Denman, Sridha Sridharan, Clinton Fookes, "Soft + Hardwired attention: An LSTM framework for human trajectory prediction and abnormal event detection". Neural Networks, Vol. 108, pp. 466-478, December 2018, https://doi.org/10.1016/j.neunet.2018.09.002

[14] Zhixian Chen, Chao Song, Yuanyuan Yang, Baoliang Zhao, Ying Hu, Shoubin Liu and Jianwei Zhan, "Robot Navigation Based on Human Trajectory Prediction and Multiple Travel Modes". Appl. Sci. Vol. 8, 2205, 2018, doi:10.3390/app8112205
[15] Zhaoyuan Yu, Linwang Yuan, Wen Luo, Linyao Feng and Guonian Lv, "Spatio-Temporal Constrained Human Trajectory Generation from the PIR Motion Detector Sensor Network Data: A Geometric Algebra Approach". Sensors, Vol. 16, 43, 2016, doi:10.3390/s16010043
[16] Daksh Varshneya and G. Srinivasaraghavan, "Human Trajectory Prediction using Spatially aware Deep Attention Models". arXiv:1705.09436v1 [cs.LG] 26 May 2017.
[17] Zhao Pei, Xiaoning Qi, Yanning Zhang, Miao Ma, Yee-Hong Yang, "Human trajectory prediction in crowded scene using social-affinity Long Short-Term Memory". Pattern Recognition, Vol. 93, pp. 273-282, September 2019, https://doi.org/10.1016/j.patcog.2019.04.025
[18] G. Habibi and J. P. How, "Human Trajectory Prediction Using Similarity-Based Multi-Model Fusion," in IEEE Robotics and Automation Letters, vol. 6, no. 2, pp. 715-722, April 2021, doi: 10.1109/LRA.2020.3048652.
[19] Palwasha Afsar, Paulo Cortez, Henrique Santos, "Automatic human trajectory destination prediction from video". Expert Systems with Applications, Vol. 110, pp. 41-51, 15 November 2018, https://doi.org/10.1016/j.eswa.2018.03.035
[20] Y. Ji, L. Wang, W. Wu, H. Shao and Y. Feng, "A Method for LSTM-Based Trajectory Modeling and Abnormal Trajectory Detection," in IEEE Access, vol. 8, pp. 104063-104073, 2020, doi: 10.1109/ACCESS.2020.2997967.
[21] Y. Qi, C. B. Soh, E. Gunawan, K. -S. Low and R. Thomas, "Assessment of Foot Trajectory for Human Gait Phase Detection Using Wireless Ultrasonic Sensor Network," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 24, no. 1, pp. 88-97, Jan. 2016, doi: 10.1109/TNSRE.2015.2409123.
[22] P. Kothari, S. Kreiss and A. Alahi, "Human Trajectory Forecasting in Crowds: A Deep Learning Perspective," in IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 7, pp. 7386-7400, July 2022, doi: 10.1109/TITS.2021.3069362.
[23] K. Han and S. Hong, "Detection and Localization of Multiple Humans Based on Curve Length of I/Q Signal Trajectory Using MIMO FMCW Radar," in IEEE Microwave and Wireless Components Letters, vol. 31, no. 4, pp. 413-416, April 2021, doi: 10.1109/LMWC.2021.3057867.
[24] X. Zhao, Y. Rao, J. Cai and W. Ma, "Abnormal Trajectory Detection Based on a Sparse Subgraph," in IEEE Access, vol. 8, pp. 29987-30000, 2020, doi: 10.1109/ACCESS.2020.2972299.
[25] J. Wang et al., "Anomalous Trajectory Detection and Classification Based on Difference and Intersection Set Distance," in IEEE Transactions on Vehicular Technology, vol. 69, no. 3, pp. 2487-2500, March 2020, doi: 10.1109/TVT.2020.2967865.
[26] Kusrini, Muhammad Resa Arif Yudianto and Hanif Al Fatta, "The effect of Gaussian filter and data preprocessing on the classification of Punakawan puppet images with the convolutional neural network algorithm", International Journal of Electrical and Computer Engineering (IJECE), Vol. 12, No. 4, August 2022, pp. 3752-3761, ISSN: 2088-8708, doi: 10.11591/ijece.v12i4.pp3752-3761
[27] M.N Tondra and Ebrahim Karami, "Vehicle detection and tracking", July 2021, doi

:10.13140/RG.2.2.15030.22085
[28] Atsushi Kawasaki, Dao Huu Hung and Hideo Saito, "Human trajectory tracking using a single omnidirectional camera", Conference: Irish Machine Vision and Image Processing, August 2014.
[29] Isaack Adidas Kamanga, "Improved Edge Detection Using Variable Thresholding Technique and Convolution of Gabor With Gaussian Filters", Signal & Image Processing: An International Journal (SIPIJ) Vol.13, No.5, October 2022, doi: 10.5121/sipij.2022.13501
[30] Padira S. V. V. N. Chanukya & T. K. Thivakaran, "Multimodal biometric cryptosystem for human authentication using fingerprint and ear", Multimedia Tools and Applications, 7 september 2019, doi :10.1007/s11042-019-08123-w
[31] Shufu Xie, Shiguang Shan, Xilin Chen and Jie Chen, "Fusing Local Patterns of Gabor Magnitude and Phase for Face Recognition", IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 19, NO. 5, MAY 2010
[32] Vishweshwrayya C. HAllur and R.S. Hegadi, "Handwritten Kannada numerlas recognition esing deep learning convolutional neural network (DCNN) classifier", CSIT, 2020.
[33] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin, "Attention Is All You Need", 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA
[34] K. Vijayaprabakaran and K. Sathiyamurthy, "Towards activation function search for long short-term model network: A differential evolution based approach", Journal of King Saud University – Computer and Information Sciences 34 (2022) pp: 2637–2650, doi : 10.1016/j.jksuci.2020.04.015
[35] Rasoul Ameri, Chung-Chian Hsu, Shahab S. Band, Mazdak Zamani, Chi-Min Shu and Sajad Khorsandroo, "Forecasting PM 2.5 concentration based on integrating of CEEMDAN decomposition method with SVM and LSTM", Ecotoxicology and Environmental Safety 266 (2023) 115572, doi : 10.1016/j.ecoenv.2023.115572
[36] Mohammad Dehghani, Zeinab Montazeri, Eva Trojovská and Pavel Trojovský, "Coati Optimization Algorithm: A new bio-inspired metaheuristic algorithm for solving optimization problems", Knowledge-Based Systems 259 (2023) 110011, doi : 10.1016/j.knosys.2022.110011
[37] Changting Zhonga, Gang Lia and Zeng Meng, "Beluga whale optimization: A novel nature-inspired metaheuristic algorithm", Knowledge-Based Systems · September 2022, doi : 10.1016/j.knosys.2022.109215
[38] Amin Manafi Soltan Ahmadi and Samaneh Hoseini Semnani, "Human trajectory prediction using LSTM with Attention mechanism", September 2023, LicenseCC BY 4.0
[39] Mengyang Huang, Menggang Zhu, Yunpeng Xiao and Yanbing Liu, "Bayonet-corpus: a trajectory prediction method based on bayonet context and bidirectional GRU", Digital Communications and Networks 7 (2021) pp:72–81, doi : 10.1016/j.dcan.2020.03.002
[40] https://motchallenge.net/data/MOT17Det/

Nomenclature

Abbreviation	Description
ADE	Average Displacement Error
ANN	Artificial Neural Network

Bi-GRU	Bidirectional Gated Recurrent Unit
Bi-LSTM	Bidirectional Long Short-Term Memory
BWO	Beluga Whale Optimization
CNN	Convolutional Neural Network
COA	Coati Optimization Algorithm
DL	Deep Learning
DNN	Deep Neural Network
DRDT	Dynamic Range-Doppler Trajectory
FDE	Final Displacement Error
FMCW	Frequency-Modulated Continuous-Wave
GNSS	Global Navigation Satellite System
HOG	History Of Oriented Gradients
HSR	Human Support Robot
IDCNN	Deep Convolutional Neural Network
LGXOR	Local Gabor XOR Pattern
LGXP	Local Gabor Xor Pattern
LSTM	Long Short-Term Memory
ML	Machine Learning
MSE	Mean Squared Error
NN	Neural Network
Oxford-IHM	Oxford Indoor Human Motion
RBF	Radial Basis Function
RCS	Radar Cross-Section
RNN	Recurrent Neural Network
Social-PEC	Social Patterns Extraction Convolution
SSALVM	Social-Scene Aware Latent Variable Model
SVM	Support Vector Machine