Global Maximal Pattern and Hybrid Optimization-Driven Deep Learning Framework for Object Detection and Recognition

Jiarui Wang, [*] **Jie Yang, Ning Ye, Yihuan Wu**

[1]College of Information Science and Engineering, Northeastern University, Shenyang 110819, China
[2]Institute of Extreme Environment Electronic System, Northeastern University, Shenyang 110819, China

Corresponding Author email: yangjie@ise.neu.edu.cn

Abstract

Object detection and recognition is the most fundamental, but challenging in the field of computer vision. Object detection identifies the presence of various individual objects in an image. However, these methods suffer from several issues due to scale variations, cluttered background and different orientations. To solve these issues, this paper presents a novel method for increasing the performance of object detection using the proposed atom-exponential weighed moving average (A-EWMA) optimization. The A-EWMA is designed by integrating atom search optimization (ASO) and Exponential Weighed Moving Average (EWMA). Initially, the input image is pre-processed and the pre-processed image is fed to the object detection module, where the object is detected using bounding box segmentation. After that, the hierarchical skeleton and the proposed global maximal pattern features are utilized for extracting the features. The hierarchical skeleton iterates itself to remove all the skeleton branches corresponding to the unimportant shape regions depending on the boundary extraction method, and the proposed global maximal pattern is the new kind of descriptor to address the short-coming of the relative maximal pattern. The resultant extracted features are employed in object recognition process for recognizing the objects. In order to facilitate effective training, the Deep Regression Neural network (Deep RNN) is employed wherein, the training of Deep RNN is performed using A-EWMA algorithm. Thus, the objects from the input images are effectively detected using the Deep RNN, which is trained using the proposed A-EWMA algorithm to boost the performance of object detection mechanism. The effectiveness of proposed A-EWMA-based DeepRNN is computed which revealed maximal accuracy of 0.891, Sensitivity of 0.872, and specificity of 0.914.

1. Introduction

As, the image data is rapidly growing in the real-world environment, the image data are now being recorded in the cloud server in order to publish it in internet. Some other researchers reported that over 1.8 billion image data is being uploaded in the social platforms, like instagram, and facebook [7]. Due to advancement of computer technology, exploitation of videos and images are inevitable. The computer vision application attained substantial growth in object detection and recognition [3]. Moreover, number of cameras is installed in the computer vision application in order to perform various tasks, like visual surveillance, and traffic monitoring. Hence, this category of image data is used in different real time applications, like object identification, visual surveillance, object detection, and object classification. To manage the image data automatically, the content of image information plays a vital role in understanding the content of digital image. In the view of computer applications, object detection is considered as the important task, which significantly increases the presentation of computer-based

applications, like license plate recognition, and object tracking [7]. In general, object localization and object detection and mainly used to identify the object from image and to localize the detected object [7]. Based on the object classification task, the object detection procedure is classified into three different categories, like feature extraction, classification, and region detection [7] [5].

Object detection is a branch of visual saliency. It is the process of segmenting and localizing the foreground objects from the real-world image. Due to the wide range of applications, like image retrieval, image resizing, object recognition, and detection, the visual saliency gained more attention in the computer technology [6]. Object detection locates and classifies the target object from video or image. When compared to the specific object detection, like vehicle detection, pedestrian, and face, the generic object detection faces various issues due to the complexity of inter class variations, like inter-class interference, occlusions, truncations, and non-grid deformations [2]. Object detection is drastically improved in the scale and performance using the growth of CNN with benchmark dataset. However, the object detector in the supervised training model requires quite large number of labeled images using the boundary box annotations. As, the bounding box is very complex to acquire, the object detector of supervised training is not feasible when it is required to perform object detection for thousands of object. Therefore, the objects are categorized based on their image labels and this data is called weakly labeled data, which is quite easier to generate either through search engine or manually. The issues in learning the image detector using the weak labels are called as weakly supervised detection (WSD) [4]. The object detection process consists of various sub activities, like pedestrian detection, skeleton detection, tracking, and face detection [7]. The key role of object detection is to localize the object instances in the given image. In the object detection model, the object recognition is designed as the classification process to generate the bounding box [8].

In the recent decades, deep convolutional neural network (CNN) is widely used in the development of artificial intelligent applications. Various neural network architectures, like ResNet [12], DenseNet [13], GoogLenet [11], and VGG [10] are developed to enhance the capability of learning on the large-scale

dataset. According to the network architectures, the object detection attained better performance and also result significant progress [8]. The automated vision system is designed in [9] to recognize and locate multiple object types more accurate. The object recognition system continuously detects and classifies the objects in the scene, as the objects undergo changes in appearance or motion. Adaptive multiple object tracking (MOT) algorithm effectively tracks the objects from the scene and make decision based on the object of interest [9]. It is easy to distinguish and recognize the objects by humans, but it is difficult for the machines to recognize and identify the objects. Hence object recognition algorithms are introduced in image processing to detect and recognize the objects more accurate [3]. Object recognition is the difficult task, as the quality of image varied from image to image [3]. Image may consist of several objects, while the object recognition system has the capacity to differentiate the objects in image. Some of the applications of object recognition are, access control system, robotics, and video surveillance [3].

The purpose of the research is to obtain an object detection and recognition technique for which the proposed A-EWMA-based Deep RNN is employed. The major contribution of this research is the detection of objects using the input images. Here, the input image is initially pre-processed and then the object is detected based on the bounding box segmentation. Once the object is detected, the feature extraction is carried out using hierarchical features and the proposed global maximal pattern. Once the features are obtained, it is given for object detection using the proposed A-EWMA-based Deep RNN, which is designed by integrating ASO in EWMA. Hence, the proposed A-EWMA-based DeepRNN renders effective accuracy through facilitating the object recognition using the input images.

The major contribution of the paper is:

- **Proposed Global maximal pattern:** The global maximal pattern is introduced based on the descriptor such that a new feature is enabled with the rotative maximal pattern in order to address the short-coming of the rotative maximal pattern.

- **Proposed A-EWMA-based Deep RNN:** A classifier, A-EWMA-based Deep RNN is proposed by modifying the training algorithm of Deep RNN using A-EWMA algorithm,

which is newly derived by combining ASO and EWMA algorithm, for the optimal tuning of weights and biases. The proposed A-EWMA-based Deep RNN is adapted for recognizing the objects so that the objects are recognized accurately.

The remaining sections of the paper is arranged as follows: Section 2 elaborates description of the object detection and recognition methods utilized in the literature and the challenges faced, which are considered as the inspiration for developing the proposed technique. The proposed method for object detection and recognition based on A-EWMA-based Deep RNN is portrayed in Section 3. The results of the proposed model with the comparative analysis are depicted in Section 4 and finally Section 5 provides the conclusion.

2. Motivations

The section deliberates the literature survey of the techniques along with the disadvantages of the methods. Here, the review of eight existing methods based on object detection. The major challenge in the object detection arises due to changing scales and orientations in the image. However, the high-resolution images brought fine details of ground objects that may generate cluttered background. Thus, these challenges of the conventional techniques stood as a motivation for devising an effective object detection model.

2.1. Literature survey

The review of eight existing methods based on object detection and recognition is illustrated.Woźniak, M. and Połap, D [1] developed detection approach by analyzing the number of clusters points in conjunction with the Convolution Neural Network (CNN) as the final classifier. This framework was utilized to determine the cluster points by combining the fuzzy logic and the graphics processing. The method can be employed as an enhanced object classifier with improved accuracy and precision, but failed to consider the size of detected objects.Zhu, Y et al.[2]presented Attention CoupleNet (ACoupleNet) for incorporating local and global information, and attention-based information to enhance the performance of detection. Initially, the cascade attention structure was designed for perceiving global scene of image and class-agnostic attention maps were generated. After that, the unique fully convolutional was introduced for coupling global structure and the

local parts of object.The method generated constant results in detecting the objects, however, failed to examine the position relation between the nonlinear coupling methods and local parts.Bhuvaneswari, R. and Subban, R [3] modelled object recognition system using the point of interest and feature selection. At first, the points of interest of image were chosen based on Derivative Kadir-Brady (DKB) detector, and then, the neighbourhood pixels were chosen for further processing. The curvelet and the gabor features were extracted followed by Support Vector Machine (SVM). The method is highly efficient in determining objects with improved computational efficiency, however, failed to process with 3D images for the object discovery and recognition. Zhang, J et al.[4] developed objectness transfer approach for object detection. Here, the knowledge was modelled using invariant features that was robust to sidtribution discrepancy between existing and new categories. The method was Applicable with weakly labelled images, but poor performance was obtained while adapting certain class of objects.

Pang, J et al. [5] presented Rotative Maximal Pattern (RMP) for fusing shape and color cues into local coloring descriptor. It was discriminative, invariant, and the complementary to the pixel-wise descriptors. This approach provided improved generalization ability with different tasks, but unable to embed the learning-based framework.Peng, H et al.[6] developed structured matrix decomposition (SMD) for formulating salient object detection as the problem of low-rank and the structured-sparse matrix decomposition. The hierarchical tree-structures sparsity-inducing norm was introduced for encoding underlying structure of image in feature space, and also the Laplacian regularization was established for enlarging the distance between the salient object representation and the background. The method yielded best performance with improved detection process, however, failed to integrate the discriminative analysis for separating the structured-sparse matrices based on regional difference.Sharma, V. and Mir, R.N [7] developed an approach for saliency detection based on faster RCNN model. The bounding box regression was employed for optimal detection with the loss function. The method showed high efficiency in detecting and recognizing the objects, but the performance of the method is affected due to variations, and complexities in background.Yao, C et

al.[8] employed coexistence discriminative features for detecting multi-class object. The region-enabled annotations were initially used as image level label for learning independent discriminative features for each and every class. After that, the co-existence relation was fused as the co-existence feature on the basis of attention mechanism. The method effectively captured the relations of coexistence with different sets of objects, but failed to incorporate deeper networks

2.2 Challenges

The challenge that has been identified by analyzing the previous object detection techniques are listed below.

- The object detection faces many issues due to the multifaceted inter-class appearance like truncations, inter-class interference, non-grid deformations and occlusions [2].

- Object detection is complicated process as the quality of images varies from image to image. As many of the images endures illumination issues, noise and orientation issues. The object recognition model should be capable to detect the objects contained in the image without considering the above inadequacies, which is major challenge [3].

- In [5], Rotative Maximal Pattern (RMP) is devised for performing object recognition and detection. The method showed encouraging performance, but possess high-dimensional feature vector which increased the storage space and computation cost.

- Many saliency models were devised for computing the saliency maps of the image to determine the salient objects. The major limitation of this method is the salient object region may consist of target objects which are mixed with the high diversity object types that limited the scalability and generalization of the method [6].

- Many object detection methods are being devised in the literature for performing object detection and recognition from the images. In recent days, the deep learning models increased the precision of object detection strategies. However, the detection and categorization of objects residing in the images are a complex task because of its varied applications. Moreover, the recognition of

hidden objects is a major issue which can severely degrade the efficiency of the system, which makes the detection modules more complicated.

3. Proposed Global maximal features and Atom-Exponential Weighed Moving Average -based Deep Convolutional Neural Network for object detection

This section illustrates the novel object detection strategy using the proposed A-EWMA algorithm. Object detection is very challenging in the satellite and aerial imagery analysis. The proposed object detection involves four different phases for detecting the object, which includes, pre-processing, object detection, feature extraction, and the object recognition. Initially, the images from the database is pre-processed and the pre-processed image is subjected to the object detection module, where the object is detected using the bounding-box segmentation. Once the object detection is done, the feature extraction is progress, for which two feature descriptors is used. One of the feature descriptors is the hierarchical skeleton [17], which iterates itself to remove all the skeleton branches corresponding to the unimportant shape regions depending on the boundary extraction method. The second descriptor is proposed global maximal pattern, which is designed based on the descriptor in [5] such that a new feature is enabled with the rotative maximal pattern in order to address the short-coming of the rotative maximal pattern. Thus, a new hand-crafted feature is developed for feature extraction process. Hence, the skeletal and the hand-crafted features is taken for the object recognition module, where the Deep recurrent neural network (Deep RNN) is employed in recognizing the objects detected from the image. For effective training of the deep RNN towards the accurate object recognition, atom-exponential weighed moving average (A-EWMA) optimization is newly developed, which is the integration of the standard atom search optimization (ASO) [15] and Exponential Weighed Moving Average (EWMA) [14]. Thus, the proposed A-EWMA-based deep RNN is employed for the effective and accurate recognition of the objects in the image. Figure 1 portrays the object detection framework using proposed A-EWMA-based Deep RNN.
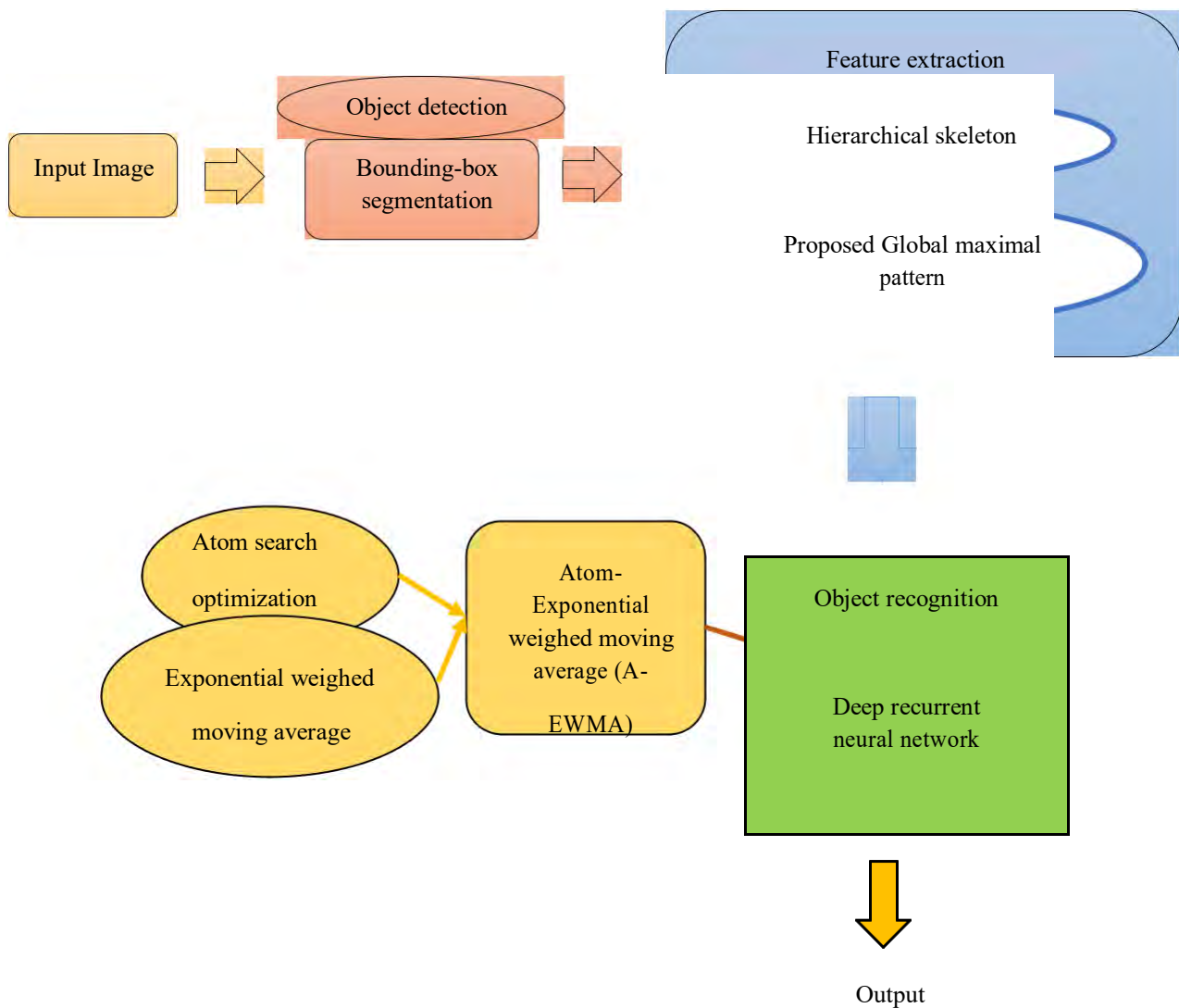
Figure 1. Schematic view of object recognition framework using the proposed A-EWMA-based Deep RNN algorithm

Assume the database B with b number of images, which is given by,

$$B = G_b \; ; \left(1 \; v \; b\right) \tag{1}$$

where, the term b represents the total images and G_b indicate the bth image. Each image G_b is subjected to the pre-processing module for better object detection. The size of the database is expressed as $U \quad V$.

3.1 Pre-processing the input image

The input image G_b is selected and is pre-processed in the pre-processing module in order to enhance the contrast of the image. The aim of pre-processing is to increase the quality of image by removing the unwanted falsification or distortions. It is required to pre-process the image in order to make the image as highly contrast and effective for detecting the object.

The result of the pre-processed image is denoted as G_b , respectively.

3.2 Object detection using Bounding Box segmentation

The pre-processed image G_b is passed into the object detection phase, where the object detection process is carried out using bounding box segmentation[18].This bounding box segmentation generates the segmented results based on several segmentation parameters. This method uses the fixed parameters to achieve inaccurate segmentation. Here, the adaptive parameter is determined based on segmentation quality assessment network. The main aim of choosing the parameter from potential parameters based on segmentation qualities of their segmented results. Figure 2 shows the block diagram of object detection approach.The steps followed in the bounding box-segmentation are described as: In the initial step, the multiple regions are segmented from image based on segmented

parameters. Then, the SQA network is utilized for measuring segmentation results and every segmentation result is assigned as score.After that, the

segmentation results are arranged using quality score for obtaining the sorted array. Here, the optimal quality is chosen as the segmentation results.
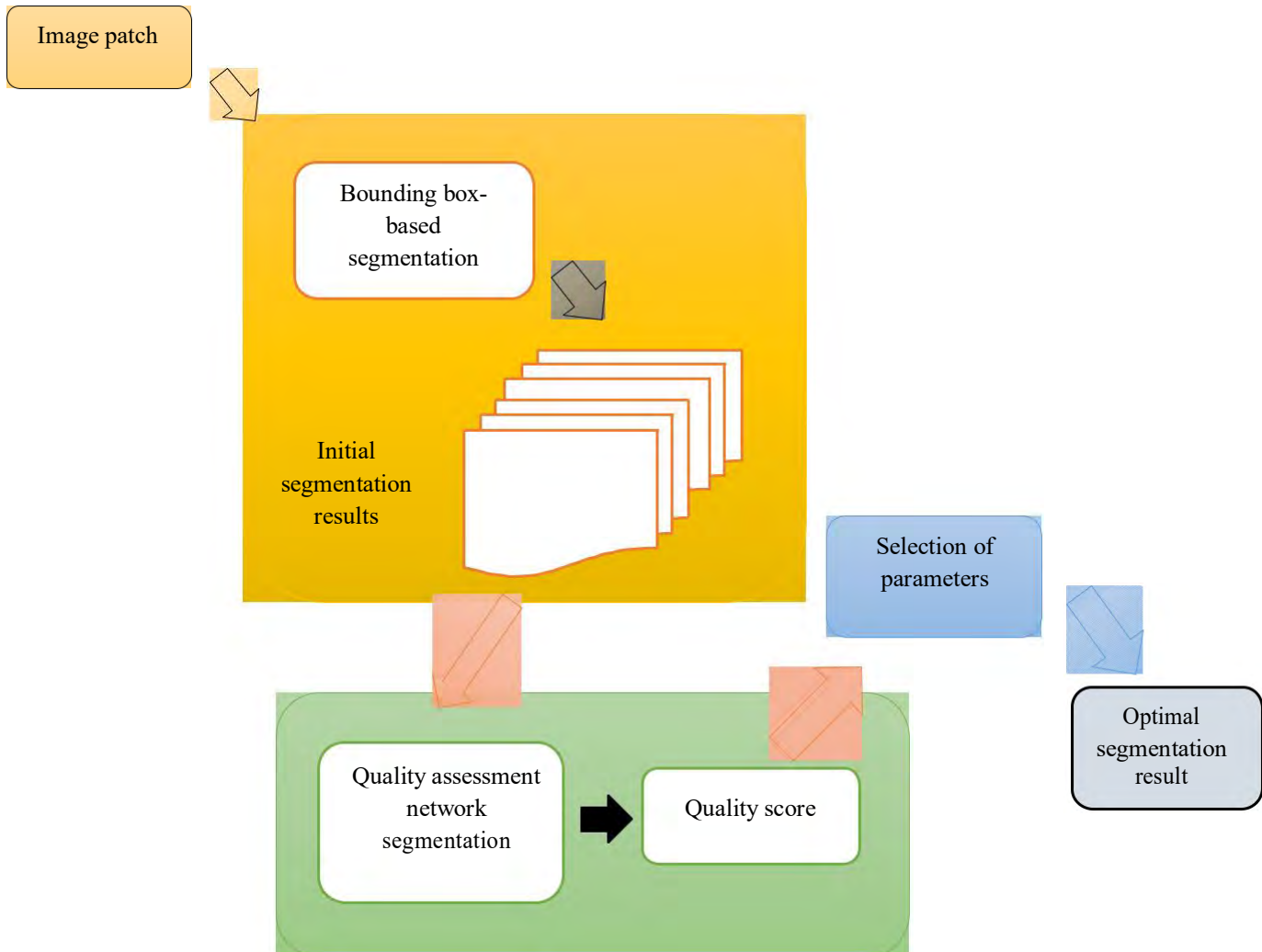
Image patch

Bounding box-based segmentation

Initial segmentation results

Selection of parameters

Quality assessment network segmentation

Quality score

Optimal segmentation result

Figure 2 Schematic view of object detection approach

a) Generation of initial segmented results

Consider the image G_b with boundary box C segments the set of segmentation results using interactive segmentation approach, and is given by,

$$T = seg\left(G_b, c, \right) \tag{2}$$

where, the bounding box-enabled segmentation is denoted as seg, and the term refer to the parameters used for the segmentation. Here, $= {}_1, ..., {}_n$ and $T = d_1, ..., d_n$ represents the set

of parameter settings and the segmentation results. The interactive Fully conventional Network (FCN) [19], and Grabcut [20] are the two kinds of bounding box segmentation approach. In Grabcut, the parameter is utilized for balancing unary and pairwise term and is

set to zero to adjust, with the value ranging from 0 and 1. The interactive FCN is obtained using boundary box as the additional input from the interactive FCN. The below steps are followed to handle the boundary box. a) The input image changed to the cropped region patch from image through boundary box. 2) The segmentation results are pretrained and finally the boundary box is slightly shifted to adjust the parameters.

b) Selection of initial result

Once the initial segmented results are generated, then the optimal one is selected as a final output to evaluate the segmentation results based on segmentation quality assessment where the selection of optimal quality is the final segmentation result. The object detection output is represented as S.

3.3 Extraction of features based on Hierarchical skeleton and the proposed Global Maximal Pattern

This section describes the extracted significant features from the input image and is very necessary for generating the highly relevant features that enable the better recognition of objects. On the other hand, the complexity of analyzing the image is minimized as the image is represented as the reduced set of features. Moreover, the accuracy associated with the object detection is assured through the effective feature extraction. Here, the object detected image S is subjected to the Hierarchical skeleton and the proposed global maximal pattern [5]for extracting the significant features, and is briefly explained below,

3.3.1 Hierarchical skeleton: The hierarchical skeleton is utilized for removing all the skeleton branches of insignificant shape regions based on boundary extraction approach. The main advantage of this method is that there is no need to search the root level when the entire hierarchical levels are arranged with pruning lines. In addition, the hierarchical skeleton uses boundary information for matching the objects by descriptor as they are hierarchically preserved. Thus, the hierarchical skeleton reduces ambiguous and increases the accuracy of single skeleton matching.

Assume the planar shape D and the initial skeleton of shape D is represented as $H^l(D)$, which is generated using max-disc model. In the circle, the centre points that are linked with the shape boundary is represented as, $e \, H^l(D)$ and the contact point of e on shape boundary are termed as generating points and the term l represents the iteration index of Discrete Curve Evolution(DCE)and is decreased till three. The term i signifies one of the following below steps.

- The Dth boundary is referred as initial polygon P^l in which the polygon simplification approach is utilized to simplify the polygon into P^i.

- Using P^i, the $H^l(D)$ is pruned to eliminate the skeletal points e $H^l(D)$ hence the generating points are remined in same contour, which is the part of shape boundary between two neighbouring vertices of P^i.

The result obtained from the individual segment is the individual pruned point in accordance with polygon partition. Thus, the un-significant points are eliminated to obtain P^{ith} simplified polygon. Consider the consecutive pair segments e_1 and e_2 that are connected with the end points e_1 e_2. Hence, similar polygon sequence, $P = P^l, P^{l-1}, ..., P^3$ so that the polygons P^{l-i} are formed by removing the single vertex h from P^{l-i+1}. The smallest shape contributed with the use of Jth measure is expressed as,

$$J(e_1, e_2) = \frac{(e_1, e_2) m(e_1) m(e_2)}{m(e_1) + m(e_2)} \quad (3)$$

where, the corner angle with the consecutive pair segments is denoted as (e_1, e_2), the length function is represented as m. When (e_1, e_2) is high, the end points $e_1 \, e_2$ is also increased using J^{th} measure. On the other side, the skeletal matching is done using the distance measure.

3.3.2 Proposed Global maximal pattern:
This subsection presents the proposed global maximal pattern, which is developed based on the descriptor in [5] such that a new feature is enabled with the rotative maximal pattern in order to address the short-coming of the rotative maximal pattern. In the global maximal pattern, the cell and its context are constructed and then, the global binary pattern are applied that results in extracted color histograms. After that, Rotatable Couple Templates (RCTs) are applied on different direction. At last, the RMP is normalize and assemble in the cell.

Conversion of images to color histograms: Initially, the image is partitioned into non-overlapping cells. For example, let us consider the pixels region 8 8 in which the color histogram is extracted. Here, the context of cell is derived in 3 3 cells. The invariance of histogram is close to color spaces. Once, the context and cells are created the global binary pattern [21] is applied to generate the set of bit strings for any direction of the binary image and these bit strings are interpreted as the binary numbers to build the global descriptor. The extracted bit strings from the intensity values is extracted on the global directed lines. In other words, the GBP is performed based on interpreting intensity values along with the direction in the image as the binary numbers, and the binary numbers are converted to the decimal values and concatenating these decimal values as elements of vector that is the

GBP shape representation. Finally, the extracted color histogram image is obtained.

Rotatable Couple Templates: In this phase, the - couple templates (-CTs) are introduced for encoding various degrees of rotations, where

belongs to $1,2,3,4$ index, which means $45, 90, 135$, and 180. Let us consider, 45 template, the local coordinate system Y_0, Y_3, and Y_4 with Y_0 as original.

Therefore, the angle between template (Y_0, Y_3) and (Y_0, Y_4) is 45.

Then, the -CTs are rotated for capturing more patterns in 3×3 context. Hence, the -CTs is extended to -f CTs, where f $1,2,...,8$ index. Naturally, -f CTs serve as the active templates

for grasping the rotation-invariant patterns.

Consider the dissimilarity function between Y_0 patch and Y_k as $e_i(M_0, M_k)$, the coupled responses in $s_1{}^{,k}$, $s_2{}^{,k}$ are calculated using the below expression,

$$s_1{}^{,k} = e_2(M_0, M_k), \ for \ k = 1,2,...,8 \qquad (4)$$

$$s_2{}^{,k} = e_2(M_0, M_i), \ for \ i = (k +)\%8 \qquad (5)$$

The dissimilarity function $e(M_o, M_i)$ is derived in Normalized Histogram Intersection (NHI) and is expressed as,

$$e(M_o, M_i) = 1 - \frac{\sum_{l=1}^{L} \min(p_o{}^l, p_i{}^L)}{\max\left(\sum_{l=1}^{L} p_o{}^l, \sum_{l=1}^{L} \min p_i{}^L\right)} \qquad (6)$$

If $e(M_o, M_i)$ is greater than the threshold, then the change of color distribution is captured at the patch level or else the noise is discovered. The max pooling is introduced for achieving the invariance, and is given by,

$$s_1{}^{,k} = \max(s_1{}^{,k}, s_2{}^{,k}) \qquad (7)$$

where, $s_1{}^{,k}, s_2{}^{,k}$ is denoted as the coupled responses in - f CTs. Here, the - f CTs is considered as the special Receptive Field (RF) with pooling operation.

Generation of RMP histogram: Let the RMP dictionary as $W^{,k} = _1{}^{.k},..., _n{}^{,k},..., _28{}^{,k}$, where, the term $_n{}^{,k}$ indicate n^{th} RCT. Hence, the nth bin is expressed as,

$$M(_n{}^{,k}) = (s_n{}^{,k}, _n{}^{,k}).s_n{}^{,k} \qquad (8)$$

where, the term represents the Dirac delta function, and is given by,

$$(y, z) = \begin{cases} 0 & for \ y \ z \\ 1 & for \ y + z \end{cases} \qquad (9)$$

Hence, the three-dimensional RMP of the image $_p$ 28 is produced, if the image is partitioned into p. To obtain the more discriminative

descriptor, the 3-dimensional RMPs are encoded further into feature vector, concatenating histograms from 2×2 cells into vector, and then normalize it with v_2 norm.

Finally, the feature vector obtained from the feature extraction step is expressed as,

$$E = e_i; 1 \ 1 \ n \qquad (10)$$

3.4 Object recognition using atom-exponential weighed moving average-based Deep Recurrent Neural Network

After the extraction of features, the object recognition is carried out based on the proposed A-EWMA for object recognition. Here, the feature vector E is given as input to the object recognition phase. The proposed A-EWMA is designed by integrating ASO and EWMA in Deep RNN to select the best weights. The structure of Deep RNN and their training process is depicted in the below subsections.

a) Architecture of DeepRNN

To detect the object effectively, the Deep RNN is employed, and the weights-biases and the weights-biases are achieved using the developed model optimally. The Deep RNN is the network architecture that contains multiple recurrent hidden layers in the layer of network hierarchy. In Deep RNN, the recurrent connection exists at the hidden layer. The Deep RNN classifier effectively operates under the varying input feature length based on the sequence of information. It uses the knowledge of previous state as input in the current prediction, and process the

iteration using the hidden state information. The recurrent feature makes the Deep RNN to be highly effective in working with the features. Due to the sequential pattern of information, Deep RNN is considered as the best classifier among traditional deep learning approaches. The architecture of Deep RNN is represented in figure 3.
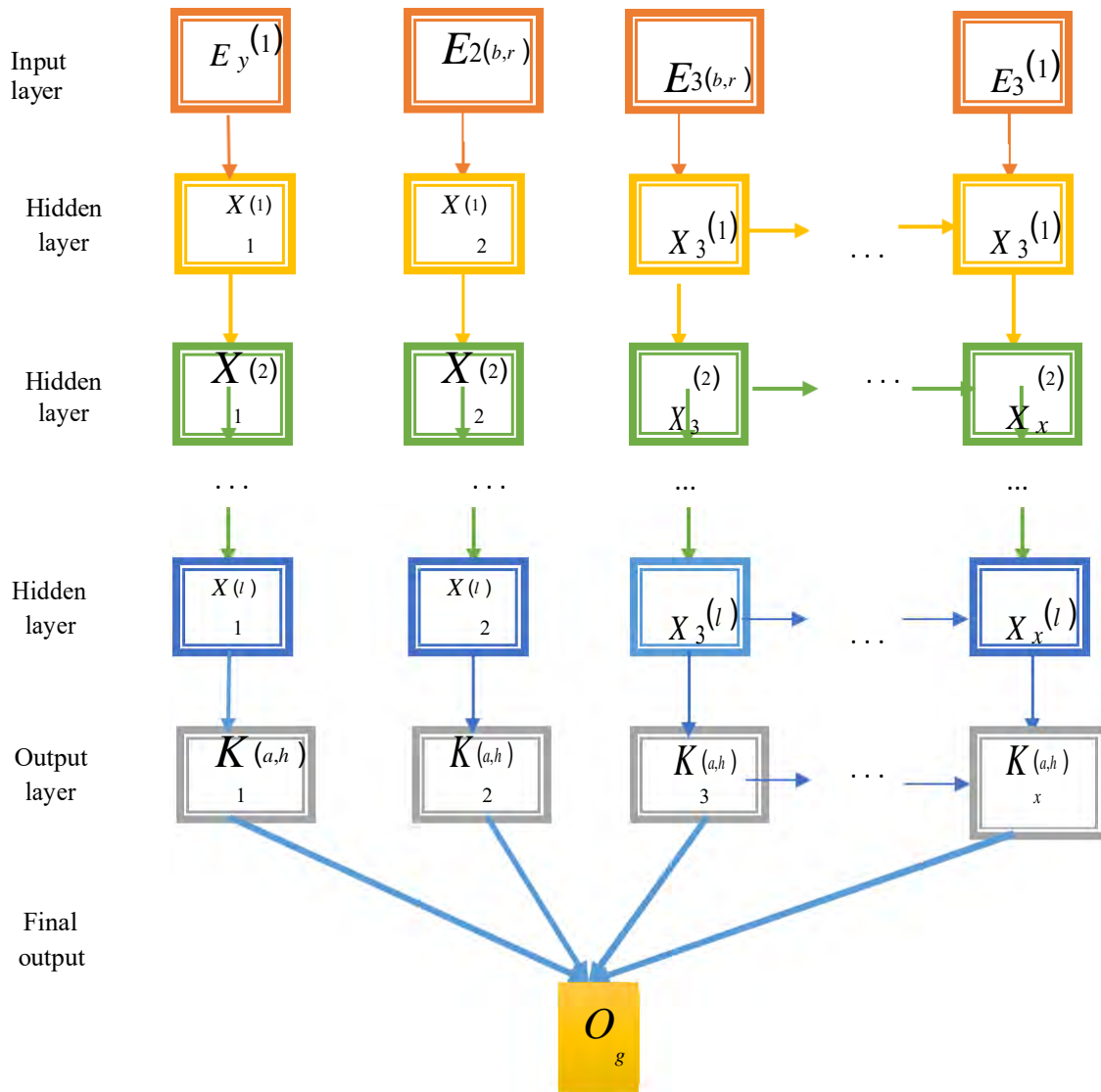

Figure 3. Architecture of Deep RNN classifier

The structure of Deep RNN is made by considering the input vector of ath layer at h^{th} time as

$$E^{(a,h)} = E_1^{(a,h)}, E_2^{(a,h)}, \dots E_j^{(a,h)}, \dots E_x^{(a,h)}$$ and the output vector of ath layer at h^{th} time as,

$$K^{(a,h)} = K_1^{(a,h)}, K_2^{(a,h)}, \dots K_j^{(a,h)}, \dots K_x^{(a,h)},$$

respectively. The pair of each elements of input and the output vectors is termed as the unit. Here, j denotes the arbitrary unit number of bth layer, and y represents the total number of units of layer. In addition to this, the arbitrary unit number and the total number of units of $(a-1)^{th}$ layer is denoted as j and C, respectively.

At this time, the input propagation weight from $(a-1)^{th}$ layer to a^{th} layer is indicated as, $I^{(a)} L^{x \, C}$, and the recurrent weight of ath layer is represented as $c^{(a)} L^{y \, y}$. Here, L refer to weights. However, the components of the input vector is expressed as,

$$E_j^{(a,h)} = q_{jy}^{(a)} \sum_{y=1}^{C} K_y^{(a-1,h)} + z_{jj}^{(a)} \sum_j K_j^{(a,h-1)} i \qquad (11)$$

where, $q_{jy}^{(a)}$ and $z_{jj}^{(a)}$ are the elements of $I^{(a)}$ and $c^{(a)}$. j denotes the arbitrary unit number of ath layer. The elements of the output vector of ath layer is represented as,

$$K_{(ja,h)} = {}_{(a)}\left(E_{(ja,h)}\right) \tag{12}$$

where, ${}_{(a)}$ denotes the activation function. However, the activation functions, like sigmoid function as $(E) = \tanh(E)$, rectified linear unit function (ReLU) as $(E) = \max(E,0)$, and the logistic sigmoid

function as $(E) = \dfrac{1}{\left(1 + e^{-E}\right)}$ are the frequently used activation function.

To simplify the process, 0th weight as $q_{(j,a_0)}$ and 0th unit as $K_0^{(a-1,h)}$ are introduced and hence the bias is represented as,

$$K_{(a,h)} = {}_{(a)}.\left(I_{(a)}K_{(a-1,h)} + C_{(a)}.K_{(a,h-1)}\right) \tag{13}$$

Here, $K^{(a,h)}$ denotes the output of classifier.

b) Training of Deep RNN using Atom-Exponential Weighted Moving Average

The training of Deep RNN is performed using the proposed A-EWMA optimization algorithm that aims at determining the optimal weights to tune the Deep RNN for recognizing the objects. The optimal weights derived from the proposed algorithm tunes the Deep RNN for deriving the optimal weights. The object recognition is progressed using the proposed A-EWMA-based Deep RNN and is effective in finding the objects through deriving the optimal weights.The proposed A-EWMA is designed by integrating ASO in EWMA, and thus attains the advantages of EWMA in ASO [15]. ASO is duly inspired by basic dynamic molecules. Here, the position of each atom lies in search space specifies the solution that is measured by the mass. All the atoms present in the population either repelorattracts each other based on the distance between them. The heavier atoms have lower acceleration, which makes them to obtain better solution in the search space. The lighter atoms have higher acceleration, which intensively makes for finding the new region in the search space. Thus, the AWOA algorithm produces easy computations and helps to understand easily. AWOA controls the exploration and exploitation phases, which render global optimal solution with higher convergence rates. The EWMA [14] is utilized for enhancing the optimization and convergence speed of Jaya. Here, the update equation of ASO is modified using the update of the EWMA. The modification makes the solution

update to be more efficient, and it further improves the convergence of the optimization algorithm. The algorithmic procedure for object recognition approach based on the developed model is given as,

Step 1: Population initialization: The first step is the population initialization of atoms in order to tackle the unconstrained optimization issues. Assume the population with C number of atoms and the position of sth atom is expressed as,

$$Z_s = w^1{}_s, \dots w_s{}^Q ; \quad s = 1, \dots o \tag{14}$$

where, $w^g{}_s \; (g = 1, \dots Q)$ refer to the cth locationin Q^{th} dimensional space of sth atom.

Step 2: Calculation of fitness function: The fitness function is evaluated based on the MSE, which is evaluated by estimated and the target outputs of the A-EWMA classifier. The fitness function is taken as the maximization function, and solution producing the maximum fitness value is considered as the optimal solution. The fitness function is given as,

$$MSE = \frac{1}{(}Y_{t \; \arg \; et} - K^{(a,h)}) \tag{15}$$

where, the number of training samples is denoted as . The terms $K^{(a,h)}$ and $Y_{t \; \arg \; et}$ refers to estimated and target output.

Step 3: Evaluation of mass: The mass of atom is measured at the lowest level on the basis of fitness value. However, the mass of atom S at iteration th is expressed as,

$$N_s() = \frac{{}_s()}{\sum\limits_{s=1}^{o} {}_s()} \tag{16}$$

where, the term $N_s()$ represents the mass, and the term ${}_s()$ is given as,

$${}_s() = \frac{W_s - W_{best}}{e^{W_{worst} - W_{best}}} \tag{17}$$

Here, the term W_{best} and W_{worst} are indicated as, $W_{best} = \min\limits_{s=1,\dots o} W_s$, and $W_{worst} = \max\limits_{s=1,\dots o} W_s$, respectively.

Step 4: Determination ofneighbors: Each atom interacts with the other atoms based on the best fitness value with neighbors. Therefore, with the function of time, F gradually minimizes with respect to the lapse of iterations. Hence F is computed as,

$$( )= o - (o - 2) \sqrt{} \tag{18}$$

Step 5:   Computation of total and constraint force:

The sum of all the components being used with the random weight that is acted on c^{th} atom from other atoms is considered as the total force and is expressed as,

$$L_s ( )= \underset{v\ Fb\ est}{rand_v O_{mv}} ( ) \tag{19}$$

where, the term $L_s ( )$ represents the total force, and $rand_v$ denotes the random number that lies in 0 to1, respectively.

Each atom in the population space is acted based on the constraint force from the best atom, and so the constraint force of sth atom is represented as,

$$_s ( )= ( )\big(Z_{b\ est} ( )- Z_s ( )\big) \tag{20}$$

where, $( )$ denotes the lagrangian multiplier, $Z_b{}^c{}_{est} ( )$ refers to the location of optimal atom at th iteration, and $Z_s ( )$ signifies the constraint force. Here, the lagrangian multiplier is represented as,

$$( )= e^{\frac{2\,0}{}} \tag{21}$$

where, indicates the multiplier weight.

Step 6:   Computation of acceleration: Based on the geometric constraint and the total force, the acceleration of sth atom at th time is expressed as,

$$T_s ( )= \frac{L_s()}{N_s ( )} + \frac{_s()}{N_s ( )} \tag{22}$$

where, $L_s ( )$ is the total force, $_s ( )$ is the constraint force, and $N_s ( )$ denotes the mass.

Step 7:   Update the velocity: Based on the $(+1)$ iteration, the velocity of sth atom is updated as,

$$_s (+1)= rand_s\ v_s ( )+ T_s ( ) \tag{23}$$

where, the term $rand_s$ represents the random number, and $T_s ( )$ indicates the acceleration.

Step 8: Update the location of atom: The update process of the proposed AWOA is carried out by modifying the position of sth atom of ASO with the sth probability of searching for the prey in WOA optimization. The position of sth atom at $(+1)^{th}$ iteration is represented as,

$$Z_s (+1)= Z_s ( )+ _s (+1) \tag{24}$$

$$Z (+1)= Z ( )+ rand _c ( )+T( ) \tag{25}$$

In the EWMA algorithm, the weighting function is applied along with the solution for the smoothing process, and hence including the exponential weighted constant to the solution improves the convergence rate of the optimization procedure. Accordingly to the EWMA,

$$Z_{sF} ( )= Z_s ( )+ (1 -)Z_{sF} (-1) \tag{26}$$

$$Z_s ( ) = \underline{1}_{}\ _F\ _F (-1) \tag{27}$$

Substitute equation (22) in (20),

$$Z_s (+1)= \frac{Z_{sF} ( )- (1 -)Z_{sF} (-1)}{} + rand\ c_s ( )_s +T( )_s \tag{28}$$

where, the term represents the weighted exponential constant ranges from $0,1$, $T_s ( )$ denotes the acceleration of first atom, and the velocity of the atom one is indicated as $c_s ( )$.

Step 9: Compute the feasibility of solution: The fitness value of each search agent is computed using the equation (10) such that the minimal error value generates optimal fitness measure and the function with the optimal result is accepted as the best solution.

Step 10: Termination: The steps 2 to 9 are repeated till the specified iteration met or the best solution is obtained. Algorithm 1 specifies the pseudo code of the developed A-EWMA.

Algorithm 1. Pseudo code of the developed model

Input: Random population $Z_s = w^1_s, \ldots w_s^Q$; $\quad\quad s = 1, \ldots o$
Output: Best position
Initialize the set of atoms with velocity v
While stopping criteria is not satisfied
Do
Fitness function computation
Computation of mass $N_s(\quad)$ using equation (16)
Determine the neighbors
Compute total and constraint force , $L_s(\quad\quad)$ and $O_s(\quad)$
Calculate $T_s(\quad)$
Update $_s(\quad+1)$
Update $Z_s(\quad+1)$ using equation (28)
End for
End while
Return the best solution

4. Results and discussion

The results and discussion made using the proposed A-EWMA-based Deep RNN is elaborated in this section.

4.1 Experimental setup

The implementation of the proposed method is carried out in the MATLAB tool using the Visual Object Classes Challenge 2012 (VOC2012) database [16], which includes 20 object classes for recognizing the objects from the realistic scenes. It is fundamentally a supervised learning problem in that a training set of labelled images is provided.

4.2 Evaluation metrics

The performance revealed by the proposed approach is evaluated using the metrics, like accuracy, specificity, and sensitivity, respectively.

*Accuracy:*It is the rate of accurate positive measure observed from the total measurements, which is represented as,

$$= \frac{R+Q}{R+Q+F+G} \quad\quad (29)$$

where, is the accuracy, R indicates true positive, Q represents true negative, F represents false positive, and G denotes false negative, respectively.

*Sensitivity:*It is the measure of true results obtained by detecting the total objects, which is expressed as,

$$= \frac{R}{R+G} \quad\quad (30)$$

where, denotes sensitivity.

*Specificity:*It is the measure of negative values resulted from the total observations, which is indicated as,

$$= \frac{Q}{Q+F} \quad\quad (31)$$

where, denotes specificity.

4.3 Experimental Results

The experimental results obtained by the proposed A-EWMA-based Deep RNN are illustrated in figure 4. The experimental results of proposed A-EWMA-based Deep RNN algorithm is done using the set of images. Figure 4a) represents the input taken from the object detection dataset, whereas figure 4b) represents the threshold image. Figure 4c) depicts the skeleton image, and the Rotative Maximal Pattern image is shown in figure 4d).
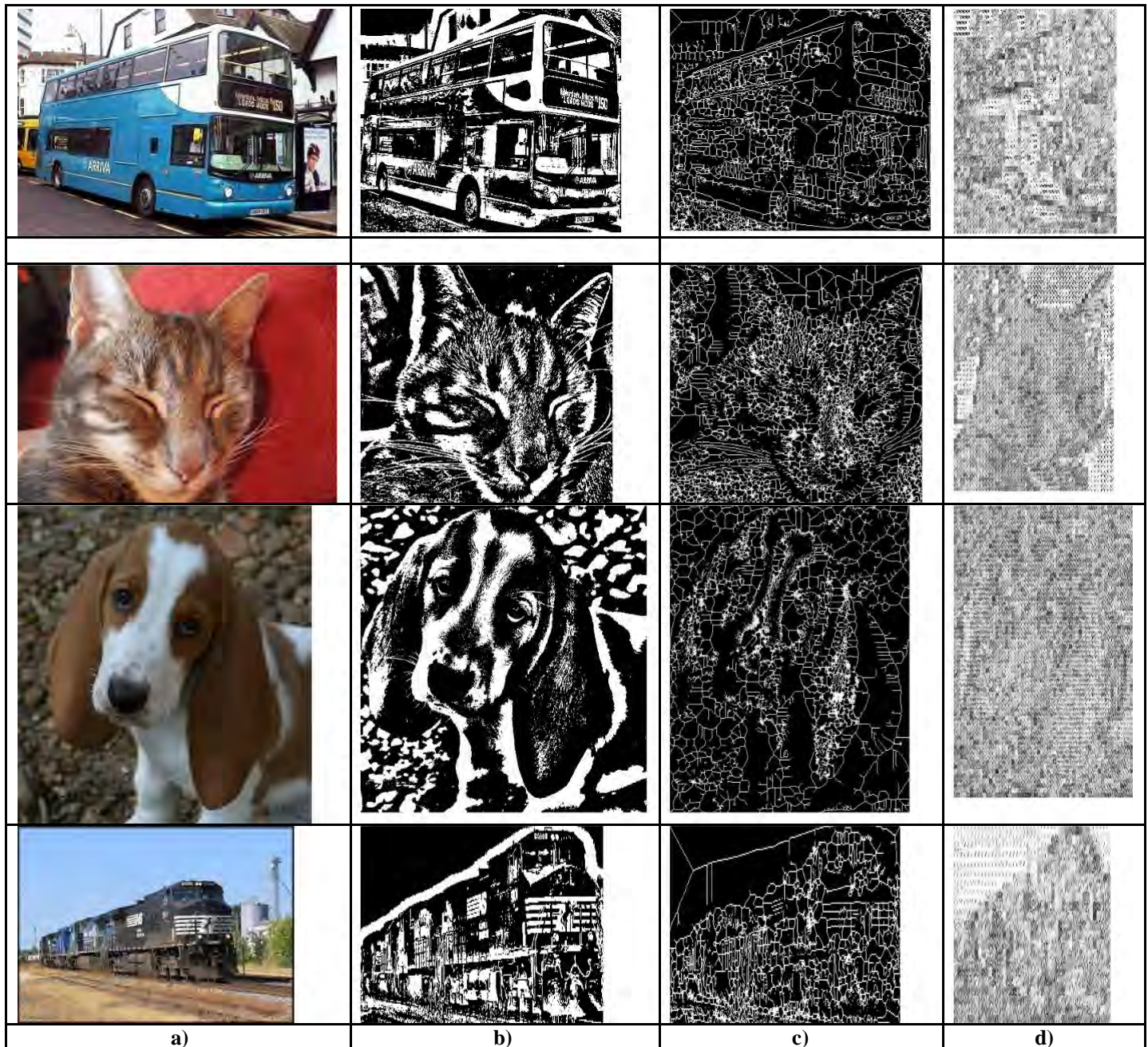
Figure 4 Sample results a) Input image, b) threshold image, c) Skeleton image, d) Rotative maximal pattern image

4.4 Performance analysis

In this section, the performance analysis of the proposed method by varying the hidden neurons, input delay, and the features is elaborately discussed.

a) Performance analysis by varying hidden neurons

Figure 5 portrays the analysis of the proposed method by varying the hidden neurons. Figure 5a) shows the performance analysis of accuracy with respect to the training data. When training data=60%, the accuracy obtained by the proposed A-EWMA-based Deep RNN with hidden neurons=5, A-EWMA-based Deep RNN with hidden neurons=10, A-EWMA-based Deep RNN with hidden neuron=15 and A-EWMA-based Deep

RNN with hidden neuron=20 is 0.708, 0.752, 0.695, and 0.657, respectively. When training data=70%, the accuracy obtained by the proposed A-EWMA-based Deep RNN with hidden neurons=5, A-EWMA-based Deep RNN with hidden neurons=10, A-EWMA-based Deep RNN with hidden neuron=15 and A-EWMA-based Deep RNN with hidden neuron=20is 0.733, 0.753, 0.697, and 0.725, respectively. When training data=80%, the accuracy obtained by the proposed A-EWMA-based Deep RNN with hidden neurons=5, A-EWMA-based Deep RNN with hidden neurons=10, A-EWMA-based Deep RNN with hidden neuron=15 and A-EWMA-based Deep RNN with hidden neuron=20 is

0.738, 0.763, 0.759, and 0.785, respectively. When training data=90%, the accuracy obtained by the proposed A-EWMA-based Deep RNN with hidden neurons=5, A-EWMA-based Deep RNN with hidden neurons=10, A-EWMA-based Deep RNN with hidden neuron=15 and A-EWMA-based Deep RNN with hidden neuron=20is 0.752, 0.773, 0.786, and 0.839, respectively.Figure 5b) represents the performance analysis of sensitivity by varying the training data. When training data=60%, the sensitivity obtained by the proposed A-EWMA-based Deep RNN with hidden neurons=5, A-EWMA-based Deep RNN with hidden neurons=10, A-EWMA-based Deep RNN with hidden neuron=15 and A-EWMA-based Deep RNN with hidden neuron=20is 0.727, 0.744, 0.710, and 0.651, respectively. When training data=70%, the sensitivity obtained by the proposed A-EWMA-based Deep RNN with hidden neurons=5, A-EWMA-based Deep RNN with hidden neurons=10, A-EWMA-based Deep RNN

with hidden neuron=15 and A-EWMA-based Deep RNN with hidden neuron=20is 0.748, 0.748, 0.730, and 0.743, respectively.Figure 5c) represents the performance analysis of specificity by varying the training data. When training data=60%, the specificity obtained by the proposed A-EWMA-based Deep RNN with hidden neurons=5, A-EWMA-based Deep RNN with hidden neurons=10, A-EWMA-based Deep RNN with hidden neuron=15 and A-EWMA-based Deep RNN with hidden neuron=20 is 0.734, 0.740, 0.719, and 0.677, respectively. When training data=70%, the specificity obtained by the proposed the accuracy obtained by the proposed A-EWMA-based Deep RNN with hidden neurons=5, A-EWMA-based Deep RNN with hidden neurons=10, A-EWMA-based Deep RNN with hidden neuron=15 and A-EWMA-based Deep RNN with hidden neuron=20is 0.755, 0.747, 0.735, and 0.761, respectively.
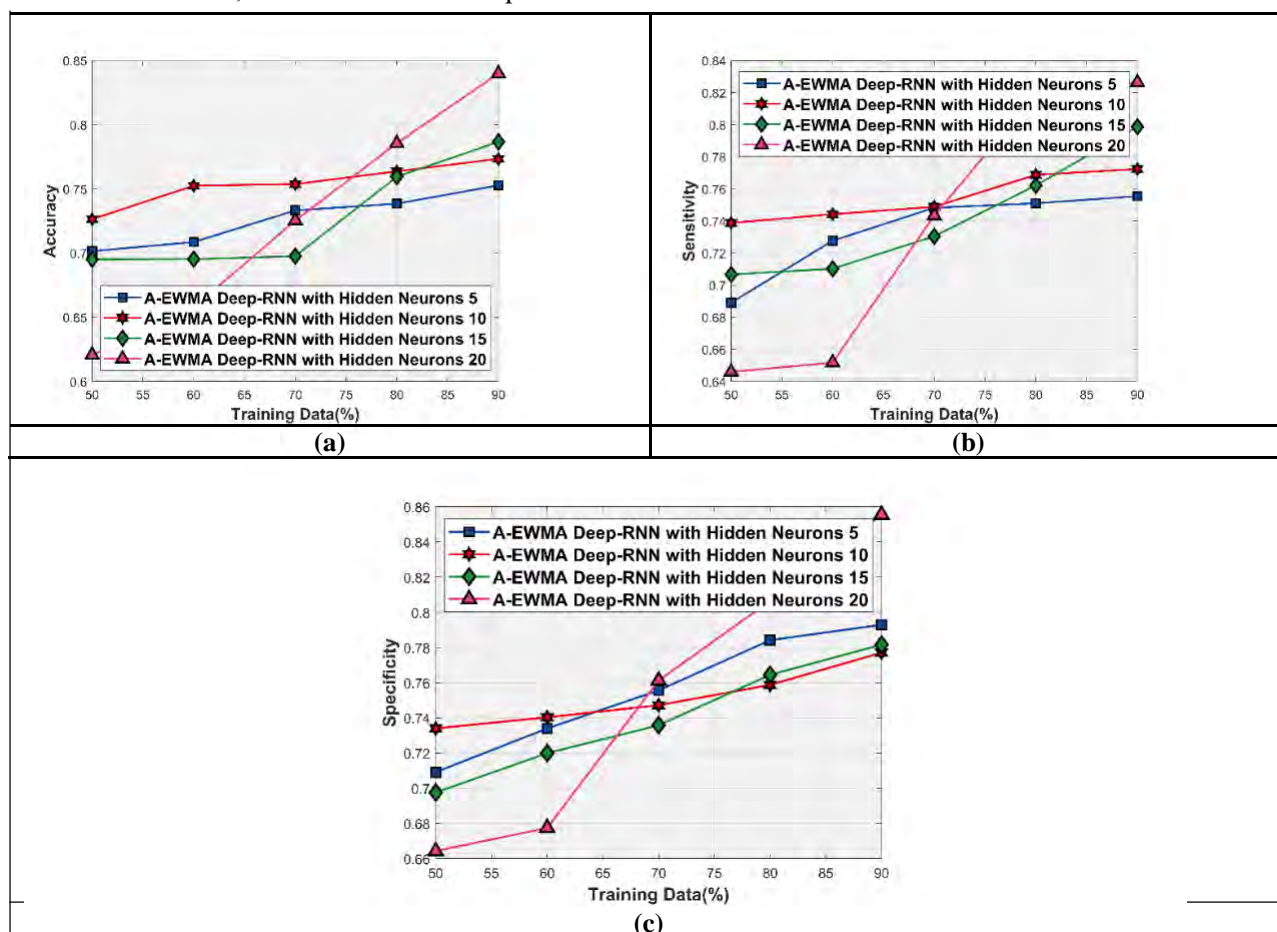

Figure 5. Performance analysis by varying hidden neurons, a) accuracy, b) sensitivity, c) specificity

b) Performance analysis by varying input delay

Figure 6 portrays the analysis of the proposed method by varying the input delay. Figure 6a) shows the performance analysis of accuracy with respect to the training data. When training data=60%, the accuracy

obtained by the proposed A-EWMA-based Deep RNN with input delay 2, A-EWMA-based Deep RNN with input delay 3, A-EWMA-based Deep RNN with input delay 4, and A-EWMA-based Deep RNN with input delay 5 is 0.829, 0.820, 0.817, and 0.814,

respectively.When training data=70%, the accuracy obtained by the proposed A-EWMA-based Deep RNN with input delay 2, A-EWMA-based Deep RNN with input delay 3, A-EWMA-based Deep RNN with input delay 4, and A-EWMA-based Deep RNN with input delay 5 is 0.835, 0.831, 0.828, and 0.815, respectively. When training data=80%, the accuracy obtained by the proposed A-EWMA-based Deep RNN with input delay 2, A-EWMA-based Deep RNN with input delay 3, A-EWMA-based Deep RNN with input delay 4, and A-EWMA-based Deep RNN with input delay 5 is 0.850, 0.839, 0.836, and 0.832, respectively. When training data=90%, the accuracy obtained by the proposed A-EWMA-based Deep RNN with input delay 2, A-EWMA-based Deep RNN with input delay 3, A-EWMA-based Deep RNN with input delay 4, and A-EWMA-based Deep RNN with input delay 5 is 0.854, 0.851, 0.838, and 0.834, respectively.Figure 6b) represents the performance analysis of sensitivity by varying the input delay. When training data=60%, the sensitivity obtained by the proposed A-EWMA-based Deep RNN with input delay 2, A-EWMA-based Deep RNN with input delay 3, A-EWMA-based Deep RNN

with input delay 4, and A-EWMA-based Deep RNN with input delay 5 is 0.805, 0.802, 0.781, and 0.772, respectively. When training data=70%, the sensitivity obtained by the proposed A-EWMA-based Deep RNN with input delay 2, A-EWMA-based Deep RNN with input delay 3, A-EWMA-based Deep RNN with input delay 4, and A-EWMA-based Deep RNN with input delay 5 is 0.809, 0.803, 0.782, and 0.775, respectively.Figure 6c) represents the performance analysis of specificity by varying the input delay. When training data=60%, the specificity obtained by the proposed A-EWMA-based Deep RNN with input delay 2, A-EWMA-based Deep RNN with input delay 3, A-EWMA-based Deep RNN with input delay 4, and A-EWMA-based Deep RNN with input delay 5 is 0.913, 0.908, 0.906, and 0.895, respectively. When training data=70%, the specificity obtained by the proposed A-EWMA-based Deep RNN with input delay 2, A-EWMA-based Deep RNN with input delay 3, A-EWMA-based Deep RNN with input delay 4, and A-EWMA-based Deep RNN with input delay 5 is 0.921, 0.931, 0.928, and 0.927, respectively.
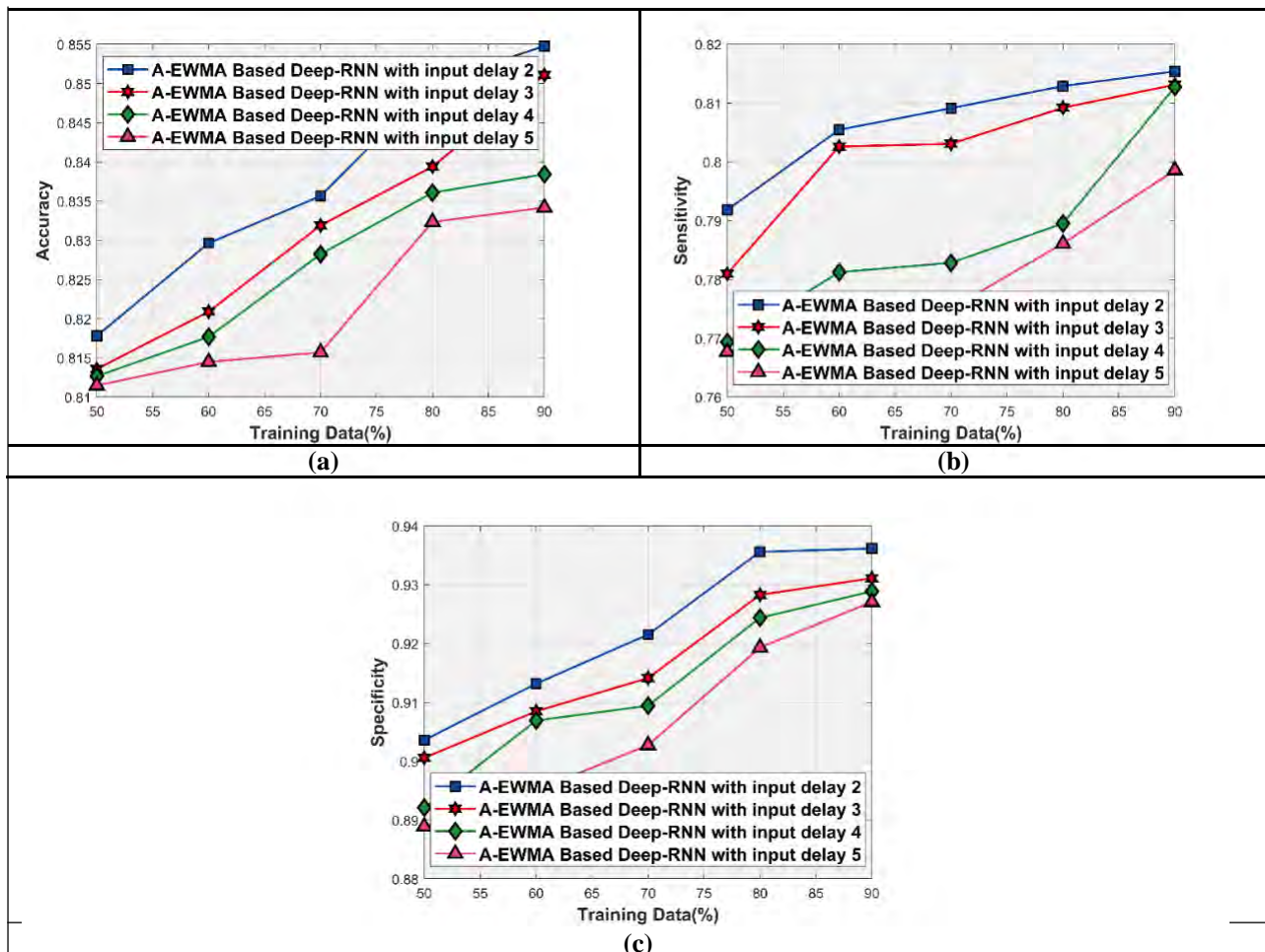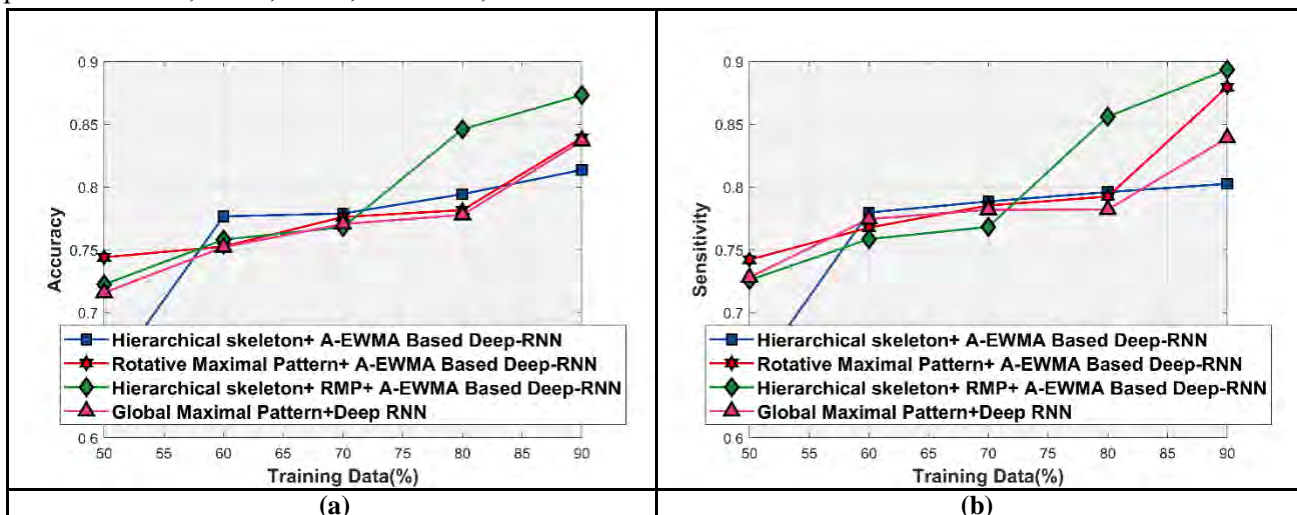

Figure 6. Performance analysis by varying input delay, a) accuracy, b) sensitivity, c) specificity

c) Performance analysis by varying features

Figure 7 portrays the analysis of the proposed method by varying the features. Figure 7a) shows the performance analysis of accuracy with respect to the training data. When training data=60%, the accuracy obtained by the Hierarchical skeleton+ A-EWMA-based Deep RNN, Rotative Maximal Pattern+ A-EWMA-based Deep RNN, Hierarchical Skeleton+ RMP+A-EWMA-based Deep RNN, and the proposed Global Maximal Pattern+ Deep RNN is 0.776, 0.752, 0.758, and 0.752, respectively. When training data=70%, the accuracy obtained by the Hierarchical skeleton+ A-EWMA-based Deep RNN, Rotative Maximal Pattern+ A-EWMA-based Deep RNN, Hierarchical Skeleton+ RMP+A-EWMA-based Deep RNN, and the proposed Global Maximal Pattern+ Deep RNN is 0.778, 0.776, 0.768, and 0.770, respectively. When training data=80%, the accuracy obtained by the Hierarchical skeleton+ A-EWMA-based Deep RNN, Rotative Maximal Pattern+ A-EWMA-based Deep RNN, Hierarchical Skeleton+ RMP+A-EWMA-based Deep RNN, and the proposed Global Maximal Pattern+ Deep RNN is 0.794, 0.781, 0.846, and 0.777, respectively. When training data=90%, the accuracy obtained by Hierarchical skeleton+ A-EWMA-based Deep RNN, Rotative Maximal Pattern+ A-EWMA-based Deep RNN, Hierarchical Skeleton+ RMP+A-EWMA-based Deep RNN, and the proposed Global Maximal Pattern+ Deep RNN is 0.813, 0.839, 0.873, and 0.836,

respectively.Figure 7b) represents the performance analysis of sensitivity by varying the features. When training data=60%, the sensitivity obtained by the Hierarchical skeleton+ A-EWMA-based Deep RNN, Rotative Maximal Pattern+ A-EWMA-based Deep RNN, Hierarchical Skeleton+ RMP+A-EWMA-based Deep RNN, and the proposed Global Maximal Pattern+ Deep RNN is 0.779, 0.767, 0.758, and 0.774, respectively. When training data=70%, the sensitivity obtained by the Hierarchical skeleton+ A-EWMA-based Deep RNN, Rotative Maximal Pattern+ A-EWMA-based Deep RNN, Hierarchical Skeleton+ RMP+A-EWMA-based Deep RNN, and the proposed Global Maximal Pattern+ Deep RNN is 0.788, 0.785, 0.768, and 0.781, respectively.Figure 7c) represents the performance analysis of specificity by varying the features. When training data=60%, the specificity obtained by the Hierarchical skeleton+ A-EWMA-based Deep RNN, Rotative Maximal Pattern+ A-EWMA-based Deep RNN, Hierarchical Skeleton+ RMP+A-EWMA-based Deep RNN, and the proposed Global Maximal Pattern+ Deep RNN is 0.772, 0.726, 0.724, and 0.749, respectively. When training data=90%, the specificity obtained by the Hierarchical skeleton+ A-EWMA-based Deep RNN, Rotative Maximal Pattern+ A-EWMA-based Deep RNN, Hierarchical Skeleton+ RMP+A-EWMA-based Deep RNN, and the proposed Global Maximal Pattern+ Deep RNN is 0.782, 0.866, 0.872, and 0.831, respectively.
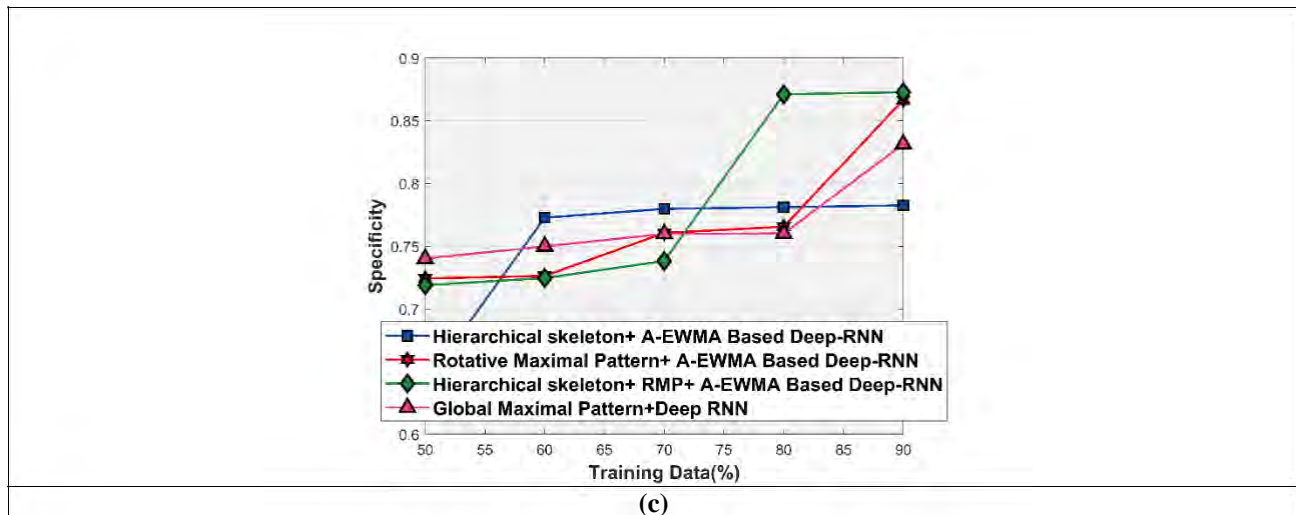

(a)

(b)

(c)

Figure 7. Performance analysis by varying the features, a) accuracy, b) sensitivity, c) specificity

4.4 Comparative methods

The performance increased by the proposed method is evaluated by comparing the proposed with the existing methods, like Fully Convolutional Neural Network [2], Support Vector Machine [3], RMP [5]+ Deep RNN respectively.

4.5 Comparative analysis

Figure 8 shows the comparative of the proposed A-EWMA-based Deep RNN in terms of sensitivity, specificity, and accuracy with respect to the training data. Figure 6 a) represents the comparative analysis of the proposed A-EWMA-based Deep RNN in terms of accuracy. When training data=60%, the accuracy obtained by the existing Fully CNN, SVM, and RMP + Deep RNN is 0.778, 0.741, and 0.749, while the proposed A-EWMA-based Deep RNN obtained better accuracy of 0.828, respectively. When training data=70%, the accuracy obtained by the proposed A-EWMA-based Deep RNN is 0.851, while the percentage of improvement reported by the proposed method in comparison with the exiting Fully CNN, SVM, and RMP + Deep RNN is 0.781, 0.741, and 0.851, respectively.Figure 8b) represents the comparative analysis of the proposed A-EWMA-based Deep RNN in terms of sensitivity by varying the training data. When training data=60%, the sensitivity obtained by the existing Fully CNN, SVM, and RMP + Deep RNNis 0.746, 0.715, and 0.727, while the proposed A-EWMA-based Deep RNN obtained better

sensitivity of 0.835, respectively. When training data=70%, the sensitivity obtained by the proposed A-EWMA-based Deep RNN is 0.849, while the percentage of improvement reported by the proposed method in comparison with the exiting Fully CNN, SVM, and RMP + Deep RNN is 0.773, 0.723, and 0.739, respectively. When training data=80%, the sensitivity obtained by the proposed A-EWMA-based Deep RNN is 0.862, while the percentage of improvement reported by the proposed method in comparison with the exiting Fully CNN, SVM, and RMP + Deep RNN is 0.778, 0.728, and 0.747, respectively.Figure 8c) represents the comparative analysis of the proposed A-EWMA-based Deep RNN in terms of specificity by varying the training data. When training data=60%, the sensitivity obtained by the existing Fully CNN, SVM, and RMP + Deep RNNis 0.784, 0.741, and 0.760, while the proposed A-EWMA-based Deep RNN obtained better sensitivity of 0.825, respectively. When training data=70%, the sensitivity obtained by the existing Fully CNN, SVM, and RMP + Deep RNN is 0.788, 0.751, and 0.769, while the proposed A-EWMA-based Deep RNN obtained better sensitivity of 0.856, respectively. When training data=90%, the sensitivity obtained by the existing Fully CNN, SVM, and RMP + Deep RNN is 0.808, 0.763, and 0.774, while the proposed A-EWMA-based Deep RNN obtained better sensitivity of 0.914, respectively.
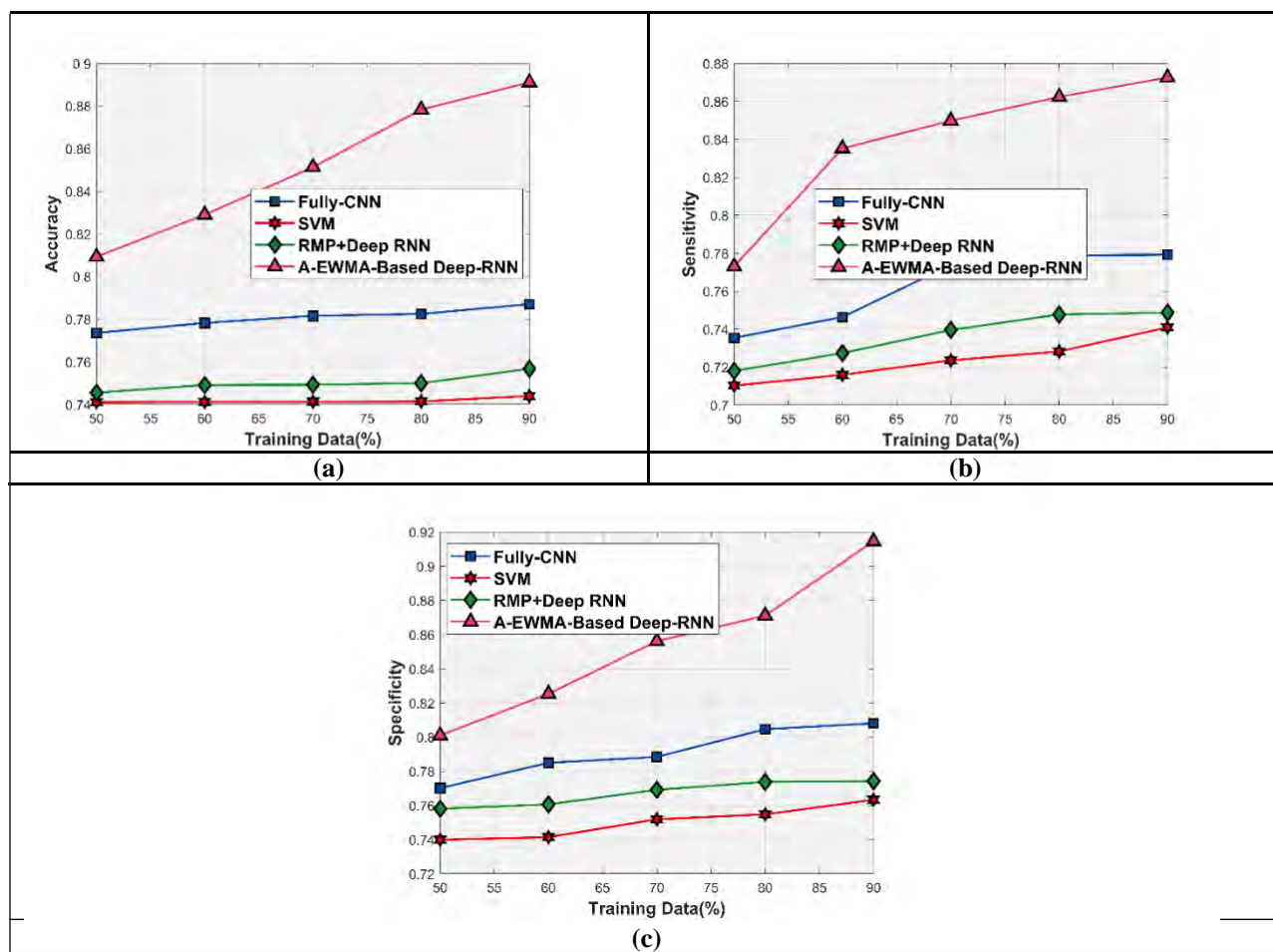
Figure 8. Comparative analysis of the developed method, a) accuracy, b) sensitivity, c) specificity

4.6 Comparative discussion

Table 1 depicts the comparative discussion of the existing Fully CNN, SVM, and RMP + Deep RNN and the proposed A-EWMA-based Deep RNN in terms of accuracy, sensitivity, and the specificityparameters by varying the training data percentage. The maximum performance measured by proposed A-EWMA-based Deep RNNin terms of accuracy parameter is 0.891, whereas the accuracy values of existing Fully CNN, SVM, and RMP + Deep RNNare 0.787, 0.743, and

0.756, respectively. The maximal sensitivity achieved by the proposed A-EWMA-based Deep RNN with value of 0.872, whereas the existing CNN, SVM, and RMP + Deep RNNacquired the sensitivity of 0.779, 0.740, and 0.872, respectively. The maximal specificity value achieved by the proposed A-EWMA-based Deep RNN is 0.914, whereas the specificity values of existing Fully CNN, SVM, and RMP + Deep RNN are 0.808, 0.763, and 0.774.

Table 1 Comparative discussion

Methods	Accuracy	Sensitivity	Specificity
Fully CNN	0.787	0.779	0.808
SVM	0.743	0.740	0.763
RMP + Deep RNN	0.756	0.748	0.774
Proposed A-EWMA-based Deep RNN	**0.891**	**0.872**	**0.914**

5. Conclusion

This research paper introduces an approach for object recognition approach based on the proposed A-EWMA-based Deep RNN. At first, the pre-processing is performed from the input image for removing the noise in the image. Then, the pre-processing output is given to the object detection approach. Here, the object detection is done based on the bounding box segmentation for finding the objects. After the detection of objects, the feature extraction is performed based on hierarchical features and the proposed global maximal pattern in order to obtain the best features for object recognition. The features are effectively extracted and increases the efficiency of detection rate. Once the feature extraction is done, the object recognition is done based on the extracted features using Deep RNN. Here, the Deep RNN is trained by A-EWMA optimization algorithm. The A-EWMA optimization inherits the characteristic features from ASO and EWMA and effectively updates the weight of the classifier. The proposed model outperforms the existing object recognition techniques with respect to the performance of extraction result. The experimentation is carried out using Visual Object Classes Challenge 2012 (VOC2012) database. The performance of the object recognition based on developed model is computed based on accuracy, sensitivity, and the specificity. The proposed model produces the maximal accuracy of 0.891, maximal sensitivity of 0.872, and the maximal specificity of 0.914 with respect to training data percentage. In future, the research can be further extended by incorporating advanced methods for object recognition.

References

[1] Woźniak, M. and Połap, D., "Object detection and recognition via clustered features," Neurocomputing, vol.320, pp.76-84, 2018.

[2] Zhu, Y., Zhao, C., Guo, H., Wang, J., Zhao, X. and Lu, H., "Attention couplenet: Fully convolutional attention coupling network for object detection," IEEE Transactions on Image Processing, vol.28, no.1, pp.113-126, 2018.

[3] Bhuvaneswari, R. and Subban, R., "Novel object detection and recognition system based on points of interest selection and SVM classification," Cognitive Systems Research, vol.52, pp.985-994, 2018.

[4] Zhang, J., Huang, K. and Zhang, J., "Mixed supervised object detection with robust objectness transfer," IEEE transactions on pattern analysis and machine intelligence, vol.41, no.3, pp.639-653, 2018.

[5] Pang, J., Huang, J., Qin, L., Zhang, W., Qing, L., Huang, Q. and Yin, B., "Rotative maximal pattern: A local coloring descriptor for object classification and recognition," Information Sciences, vol.405, pp.190-206, 2017.

[6] Peng, H., Li, B., Ling, H., Hu, W., Xiong, W. and Maybank, S.J., "Salient object detection via structured matrix decomposition," IEEE transactions on pattern analysis and machine intelligence, vol.39, no.4, pp.818-832, 2016.

[7] Sharma, V. and Mir, R.N., "Saliency Guided Faster-RCNN (SGFr-RCNN) Model for Object Detection and Recognition," Journal of King Saud University-Computer and Information Sciences, 2019.

[8] Yao, C., Sun, P., Zhi, R. and Shen, Y., "Learning coexistence discriminative features for multi-class object detection," IEEE Access, vol.6, pp.37676-37684, 2018.

[9] Wong, S.C., Stamatescu, V., Gatt, A., Kearney, D., Lee, I. and McDonnell, M.D., "Track everything: Limiting prior knowledge in online multi-object recognition", IEEE Transactions on Image Processing, vol. 26, no. 10, pp.4669-4683, 2017.

[10] Simonyan, K. and Zisserman, A., "Very deep convolutional networks for large-scale image recognition", arXiv preprint arXiv:1409.1556, 2014.

[11] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A., "Going deeper with convolutions", In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1-9, 2015.

[12] He, K., Zhang, X., Ren, S. and Sun, J., "Deep residual learning for image recognition", In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778, 2016.

[13] Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K.Q., "Densely connected convolutional networks", In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4700-4708, 2017.

[14] Saccucci, M.S., Amin, R.W. and Lucas, J.M., "Exponentially weighted moving average control schemes with variable sampling intervals," Communications in Statistics-Simulation and Computation, vol.21, no.3, pp.627-657, 1992.

[15] Zhao, W., Wang, L. and Zhang, Z., "Atom search optimization and its application to solve a hydrogeologic parameter estimation problem," Knowledge-Based Systems, vol.163, pp.283-304, 2019.

[16] Visual Object Classes Challenge 2012 (VOC2012) database, "http://host.robots.ox.ac.uk/pascal/VOC/voc2012/ ", accessed on October 2019.

[17] Cong Yang, Oliver Tiebe, KimiakiShirahama, and Marcin Grzegorzek, "Object matching with hierarchical skeletons", Pattern Recognition, vol.55, pp.183-197, July 2016.

[18] Fanman Meng, Lili Guo, Qingbo Wu, and Hongliang Li, "A New Deep Segmentation Quality Assessment Network for Refining Bounding Box Based Segmentation", IEEE Access, vol.7, pp.59514-59523, 2019.

[19] Jonathan Long, Evan Shelhamer, and Trevor Darrell, "Fully Convolutional Networks for Semantic Segmentation", In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3431-3440, 2015.

[20] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake,"GrabCut"-Interactive Foreground Extraction using Iterated Graph Cuts", In ACM transactions on graphics (TOG), vol. 23, no. 3, pp. 309-314, August 2004.

[21] Erdal Sivri and Sinan Kalkan,"Global Binary Patterns: A Novel Shape Descriptor", In MVA, pp. 169-172, 2013.