Computer Science Review 39 (2021) 100320

Contents lists available at ScienceDirect

Computer Science Review

journal homepage: www.elsevier.com/locate/cosrev

Review article

Automatic hand gesture recognition using hybrid meta-heuristic-based feature selection and classification with Dynamic Time Warping

Manisha Kowdiki[1,*], Arti Khaparde[2]

Dr. Vishwanath Karad MIT World Peace University, School of Electronics & Communication Engineering Department, Pune, India

ARTICLE INFO

Article history:
Received 6 June 2020
Received in revised form 3 October 2020
Accepted 14 November 2020
Available online 26 November 2020

Keywords:
Hand gesture recognition
Static and dynamic type
Optimal feature selection
Optimal trained neural network
Dynamic Time Warping
Deer Hunting-Based Grey Wolf Optimization

ABSTRACT

Of late, the research world has been vigorously involved in, inventing strategy and techniques to improve the spontaneity of Human Computer Interaction (HCI). Gesture recognition is one of the most probable techniques in this area. The eventual aim here is to introduce an intelligent system for hand gesture recognition in both static and dynamic area, which is still a challenging point due to the lag of valuable beneficial methods. The main intent of this paper is to implement an efficient hand gesture recognition model considering both static and dynamic datasets for Indian Sign Languages (ISL). In static type, images are taken for processing, whereas video frames are used for processing the dynamic type. The proposed recognition model involves five main steps "(a) Image pre-processing, (b) gesture segmentation, (c) Feature extraction, (d) Optimal Feature Selection, and (e) Recognition". In the pre-processing phase, greyscale conversion and histogram equalization are performed. The pre-processed image is subjected to the segmentation process, where the Active Contour model and Canny Edge Detection is implemented. In the feature extraction phase, both the contour image, and the edge detected image is deployed, in which Histogram of Oriented Gradients (HOG) features are extracted from the contour image, and Edge Oriented Histogram (EOH) features are extracted from edge detected images. To reduce the dimension of HOG, and EOH features, Principle Component Analysis (PCA) is applied. Further, the region props features are extracted for both contour and edge detected image. Finally, all these features are summed, and the optimal feature selection process performs here to select the unique feature giving different information with less correlation. Finally, the recognition classifier called Neural Network (NN) is adopted, where the new training algorithm is used to update network weight. Dynamic Time Warping (DTW) method helps to remove the repeated frames in the video and to reduce the time consumption of testing. In both feature selection and classification, a hybrid algorithm Deer Hunting-based Grey Wolf Optimization (DH-GWO) is used for selecting the features and weight update in NN as well. Hence, the integration of a hybrid meta-heuristic algorithm is highly efficient for recognizing the characters for images and words for videos with high recognition accuracy.

Contents

* Corresponding author.
E-mail addresses: manisha.kowdiki@mitwpu.edu.in (M. Kowdiki), arti.khaparde@mitwpu.edu.in (A. Khaparde).
1 Assistant professor.
2 Head of School.

https://doi.org/10.1016/j.cosrev.2020.100320
1574-0137/

Nomenclature

HCI	Human Computer Interaction
ISL	Indian Sign Languages
HOG	Histogram of Oriented Gradients
EOH	Edge Oriented Histogram
PCA	Principle Component Analysis
NN	Neural Network
DTW	Dynamic Time Warping
GWO	Grey Wolf Optimization
DHOA	Deer Hunting Optimization Algorithm
DH-GWO	Deer Hunting-based Grey Wolf Optimization
LMC	Leap Motion Controller
CNN	Convolutional Neural Network
SDTW	Structured Dynamic Time Warping
CLT	Continuous Letter Trajectory
WBCFs	Wristband-Based Contour Features
DDNN	Deep Dynamic Neural Network
HMM	Hidden Markov Model
DBN	Deep Belief Network
ASL	American Sign Language
SVM	Support Vector Machine
FDR	False Discovery Rate
PSO	Particle Swarm Optimization
KNN	K-Nearest Neighbour
NPV	Negative Predictive Value
WOA	Whale Optimization Algorithm
FPR	False Positive Rate
MCC	Matthew's Correlation Coefficient
NB	Naive Bayes
FNR	False Negative Rate

1. Introduction

Gestures are expressive, significant body movements consisting of physical activities of hands, fingers, face, arms, body, or head to send important data or communicate with the environment [1,2]. Moreover, the hand gesture is one of the major natural, expressive, and usual kinds of body language for transferring attitudes and emotions in human associations [3,4]. By protesting the virtual reality world, authoritative tools are hugely favoured for utilizing hand gestures than remaining gesture signatures. Moreover, a hand gesture permits the usage of gestures formed to navigate and maintain within the virtual reality world. Therefore, for interpreting the hand gestures of the human inside the virtual reality world, the suitable approaches for gesture recognition are necessitated. Also, gesture recognition is used for identifying the class labels from a video or an image that comprises gestures provided by the user [5,6].

Also with virtual reality, the gesture recognition model is employed in numerous areas. Hand gestures are split into two kinds such as "static and dynamic" [7,8]. The static gesture is a signature, in which the actions of the hand are not a crucial part of the gesture. Moreover, the crucial point of static gesture in the form of the hand itself. Simultaneously, the dynamic hand gesture utilizes the activities of the hand and the shape as the important point of the gesture. This type of recognition is a decisive part of human movement recognition. However, the job is challenging due to more shape variance and serious obstruction observed among fingers [9]. It is quite complicated for seizing this type of ample dynamic hand gestures using a monocular video sensor, and these defect constraints video-based hand gesture recognition performance [10,11]. In the past decades, innovative depth sensors like Microsoft Kinect sensor and LMC [12] that produced three-dimensional depth information of the scene have contributed more to object segmentation and three-dimensional hand gesture recognition. The key components of a hand recognition model are gesture recognition, hand feature identification, data acquisition, and hand localization based on recognized features. The classic solution for data acquisition is colour cameras, which have already been successfully used for gesture recognition tasks [13–15]. These solutions are, however, lighting conditions, sensitive to clutter, and skin colour. Video capture has an additional defect associated with the action speed.

Various methods were used for hand localization in obtained information, for their nature. The traditional solution is depth thresholding to depth data whether empirical or automated. Empirical solutions select the restrictions of the major possible search space using trial and error and focus the calculation effort on hand localization within it. Based on the assumption, the

automated solutions are the detection of the nearest point to the camera [16], which the hand is the nearer object in the scene, and the usage of other reference elements in the scene like facial colour data [15] and head position [17] for recognizing the significant probable location of the hand. A complete solution is to scan the whole visibility space of the Kinect by overlapping depth intervals until the hands are seemed [12]. However, there are several defects present in [13–15]. The defects present as the following reasons such as the diversity and flexibility of human gesture that means even a similar person does a similar gesture two times, the two gestures are distinct.

- To put forward a new concept of hand gesture recognition by static and dynamic basis using different steps like Image pre-processing, gesture segmentation, Feature extraction, Optimal Feature Selection, and Recognition.
- To segment the hand gestures using the Active Contour model and Canny Edge Detection, and to extract the HOG features from a contour image, EOH features from edge detected image. Also, region props features are extracted from both contour and edge detected image.
- To perform the optimal feature selection using the proposed DH-GWO-based meta-heuristic algorithm to speed up the recognition model.
- To develop the optimal trained NN using the DH-GWO-based weight update for maximizing the recognition accuracy.
- To induce the concept of DTW into dynamic hand gesture recognition that intends to eliminate the repeated frames from videos, thus reducing the processing time.

The entire paper is organized according to the series mentioned below. Literature Review and features and challenges of existing hand gesture recognition models are specified in Section 2. Moreover, Section 3 describes the developed static and dynamic hand gesture recognition system. The procedure for gesture segmentation and feature extraction is shown in Section 4. Optimal feature selection and detection for static and dynamic gesture recognition are depicted in Section 5. The developed DH-GWO for optimal gesture recognition is described in Section 6. At last, the results and discussions are shown in Section 7. Finally, the conclusions of the paper are given in Section 8.

2. Literature review

2.1. Related works

In 2016, Plouffe and Cretu [18] had introduced a novel algorithm to enhance the time of scanning for recognizing the starting pixel on the handshape in a particular space. Moreover, a new model named the directional search model was deployed for detecting the whole shape of the hand. Later, the k-curvature algorithm was used to trace the fingertips, and dynamic time wrapping was utilized to pick gesture persons and also to identify gestures on contrasting the identified gesture with a sequence of reference gestures recorded in advance. The results have shown that the developed approach has performed well for recognizing the sign digits, and the same regarding the performance for both static and dynamic recognition of famous signs for the sign language alphabet when compared with existing methods.

In 2016, Wu et al. [19] have represented a DDNN approach to find the multimodal gesture. A semi-supervised hierarchical dynamic framework was developed to concurrent gesture segmentation and recognition based on HMM, in which depth, RGB images, and skeleton joint data were considered as the inputs to the multimodal observations. The high-level spatiotemporal representations were learning using deep neural networks suitable for input modality: a Gaussian-Bernoulli DBN to hold skeletal dynamics, and a 3DCNN to control and merge the collection of depth and RGB images, which was attained during the modelling and learning of the discharge possibilities of the HMM to deduce the series of gestures. The performance was correlated with existing hand-tuned feature-based methodologies and more learning-based techniques.

In 2017, Oyedotun and Khashman [20] have suggested deep learning for recognizing the significant differentiation of hand gestures to the complete "24 hand gestures" taken from the "Thomas Moeslund's" gesture recognition dataset. The results have biologically inspired and DNN like CNN and loaded denoising autoencoder can learn the compound hand gesture classification task by minimum error rates. The reviewed networks were trained and examined on data taken from the Thomas Moeslund database. The results were compared over the past tasks where only minute subsets of the ASL hand gestures were taken into consideration for identification.

In 2018, Hu et al. [21] have recommended a 3D separable CNN to recognize the dynamic gesture. The system has focused to create the model easily without undermining the elevated recognition accuracy so that the model was established to improve reality glasses effectively in the future. To resolve the separation operation for the unwanted gradient diffusion and enhance the performance of the network, the implementation of the layer-wise learning rate and skip connection was used. The combination of feature data was significantly upgraded by the shuffle operation. Moreover, a dynamic hand gesture library was established by 'HoloLens' that has proved the utility of the developed approach.

In 2018, Tang et al. [22] have proffered an SDTW method to constant hand trajectory recognition. Initially, a segmentation approach named automatic continuous trajectory segmentation model was introduced that united the templates and data related to velocity for identifying the start and final points in hand gesture trajectories. Later, various weights were allotted to feature successions depending on the arranged data, from the locations of corner points in the arbitrary trajectories. At last, SDTW was validated on the CLT database, and the outputs have revealed that the suggested method was robust to the differentiation of similar handwritten letter, and usually performed well over traditional techniques.

In 2018, Lee et al. [23] have developed a framework to identify compound static hand gestures with WBCFs. The suggested framework needs the person to wear a pair of black wristbands for two hands, and then hands were segmented precisely. The foremost and pointed corner of the wrist band on a person's hand was identified initially and considered as a sign for extracting the WBCF of a hand gesture, and next a general feature matching approach was introduced to acquire a recognition output. To tackle with the cases, wherein the hand region was not segmented exactly, watershed segmentation and the region merging approaches were chosen for offering modifications on the hand region segmentation. Therefore, the test outcomes have been revealed that the proposed model was utilized to identify 29 Turkish finger spelling sign hand gestures and attained the best performance.

In 2018, Li et al. [24] have developed a sparsity-driven approach of micro-Doppler testing for recognizing the dynamic hand gesture with radar sensors. Initially, the sparse depictions of the replications reflected from dynamic hand gestures were attained during the Gaussian-windowed Fourier dictionary. Next, by the orthogonal matching pursuit model, the micro-Doppler depictions of dynamic hand gestures were taken out. At last, the nearest neighbour classifier and the improved Hausdorff distance

were united to identify the dynamic hand gestures depending on the sparse micro-Doppler representations. The tests were performed with the real radar data, which has shown that the recognition accuracy provided by the suggested approach was best, based on the principal component evaluation and deep CNN with the minimum training dataset.

In 2019, Tang et al. [25] have merged image entropy as well as density clustering for using the keyframes from the video of hand gesture for feature extraction that enhanced the identification efficiency. Furthermore, a feature fusion approach was introduced for enhancing the feature representation that increases the recognition of the performance. To evaluate the proposed method in a "wild" environment, two latest datasets such as HandGesture and Action3D datasets were established. Finally, the evaluated tests exhibited that the suggested method has attained the best outcomes on "Northwestern University, Cambridge, HandGesture and Action3D hand gesture datasets".

2.2. Review

Though there are many static and dynamic modes of hand gesture recognition methodologies, many learning strategies are still in need of attaining hand gesture recognition accuracy. A few of the features and challenges are described in Table 1. Among them, the k-curvature algorithm [18] k-curvature algorithm is a low complex method, and DTW [18] is very easy to train. But, DTW requires more effort to identify the optimal time alignment path. SVM [25] are good when they did not have any idea on the data, and Works well even with unstructured data. Though, it is having some disadvantages like selecting a good kernel function is critical, and requires more training time for huge datasets. CNN [20,21] has weight sharing feature, has automatic feature extraction, has high accuracy, and it simplifies computation to a great extent without losing any information. Yet, there are some conflicts like it is computationally expensive, requires more training data, it is comparatively slow, and the hyper-parameter is non-trivial. SDTW [22] handles the sequences of various lengths and has high recognition accuracy. Though, the method is more complicated. Watershed Segmentation [23] requires less computational time, it is simple and fast. Yet, it provides excessive over segmentation. DDNN [19] learns high-level features from the data in an incremental manner, and has high efficiency. Still, it uses black box testing. The orthogonal Matching Pursuit algorithm [24] reduces the complexity, and it is flexible and has high efficiency. However, it requires more number of inner product operations. Hence, the above specified challenges are taken into consideration in upcoming researches for improving the performance in recognizing the hand gestures in a precise manner.

3. Developed static and dynamic hand gesture recognition system

3.1. Proposed architecture

Researchers in the past years have been observing the high research interest in the growth of intuitive and natural user interfaces. This kind of interface has to remain undetectable for the users, permitting them to unobtrusively associate with an application, without any requirement of the specialized and costly tool. Moreover, they have to assist a natural association and be appropriate for the user by not impressing detailed calibration processes. Accordingly, they have to satisfy their task in real time with accuracy and produce strength against background clutter. These different necessitates and their intrinsic difficulty still generate major defects for researchers. The main aim of HCI is to enhance the interaction among the users and computers by making the computer receptive to the user requirements. Moreover, HCI using a personal computer is not just restricted to mouse and keyboard interaction. Interaction among humans comes from various sensory modes such as speech, facial, body, gesture expressions. Being the ability to interact with the system naturally is becoming most significant in several fields of HCI. Since hand gestures compose a robust inter-human communication modality, they are taken into consideration and intuitive and convenient means to communicate among machines and humans, which validates the research communities interest in the growth and advancement of hand gesture techniques. One of the major significant capabilities of an effective user interface is thus its capacity for recognizing the static as well as dynamic hand gestures. The developed hand gesture recognition system is depicted in Fig. 1.

In the proposed architectural model, effective hand gesture recognition is developed with both static and dynamic datasets using ISL, which involves both static and dynamic images. The developed hand gesture recognition model consists of five steps such as (1) Image pre-processing, (2) gesture segmentation, (3) Feature extraction, (4) Optimal feature selection, and (5) Recognition. During the pre-processing phase, the images or the video frames are initially converted into greyscale images, and the contrast of the image is enhanced by histogram equalization. Further, the pre-processed image is applied to the segmentation procedure. In segmentation, the Active Contour model, and Canny edge detection algorithm is used. This approach is implemented for splitting the background and foreground regions accurately. Once the segmentation is done by an active contour model, further, canny edge detection is performed, which is an edge detection operator that employs a multi-stage algorithm for detecting a wide range of edges present in images. Once the segmentation is performed, both contour and edge detected images are taken, in which the HOG feature is extracted from contour images and the EOG feature is taken out from edge detected images. For decreasing the dimension of HOG and EOG features, a dimension reduction technique called PCA is adopted. Later, the region props features such as "major and minor axis length, area, Euler number, equivalent diameter, bounding box, eccentricity, centroid, solidicity, extent, extrema, convex area, and orientation" are extracted for both contour and edge detected images. Finally, all the extracted features are summed and then for selecting the unique feature providing distinct information with less correlation is obtained by an optimal feature selection method. Moreover, the hybridization of two meta-heuristic algorithms like DHOA and GWO termed as DH-GWO is used for optimal feature selection. Next, the recognition classifier named NN is used, in which the new training model is utilized for updating the network weight using the proposed DH-GWO. In dynamic type (videos), the time duration of both training and testing videos of specific gesture is wrapped for providing the discrete information based on distance metric with the help of the DTW approach, which is useful for eliminating the redundant or repeated frames existing in the video, and for decreasing the time utilization of testing. Finally, the output of the hand gestures recognition for images as well as video provides the output in the form of categorization of character or words.

3.2. Image description

Let HE^{I} be considered as the input hand gesture image or video for recognition. Here, the grey scale converted image is denoted as HE^{grey} and the contrast enhanced image by histogram equalization is denoted as HE^{Hist}. Moreover in gesture segmentation, the contour image is expressed as $HE^{contour}$ and the canny-based edge detected image is expressed as HE^{edge}. The combination

Table 1
Features and challenges of existing hand gesture recognition models.

Author [citation]	Methodology	Features	Challenges
Plouffe and Cretu [18]	k-curvature algorithm and Dynamic Time Warping algorithm	• K-curvature is low complex • DTW is easy to train.	• DTW requires more effort to identify the optimal time alignment path.
Tang et al. [25]	SVM	• SVM's are good when they did not have any idea about the data. • Works well even with unstructured data.	• Selecting good kernel function is critical. • Requires more training time for huge datasets.
Hu et al. [21]	CNN	• CNN has weight sharing feature. • Has automatic feature extraction.	• It is computationally expensive. • Requires more training data.
Tang et al. [22]	SDTW	• Handles the sequences of various lengths. • Has high recognition accuracy.	• More complication
Lee et al. [23]	Watershed Segmentation	• Requires less computational time. • It is simple and fast.	• Provides excessive over segmentation.
Wu et al. [19]	DDNN	• Learns high-level features from the data in an incremental manner. • It has high efficiency.	• Still uses black box testing.
Li et al. [24]	Orthogonal matching pursuit algorithm	• Reduces the complexity. • It is flexible and has high efficiency.	• Requires more number of inner product operations.
Oyedotun and Khashman [20]	CNNs	• It has high accuracy. • It simplifies computation to great extent without losing any information.	• It is comparatively slow. • Hyper-parameter tuning is non-trivial.

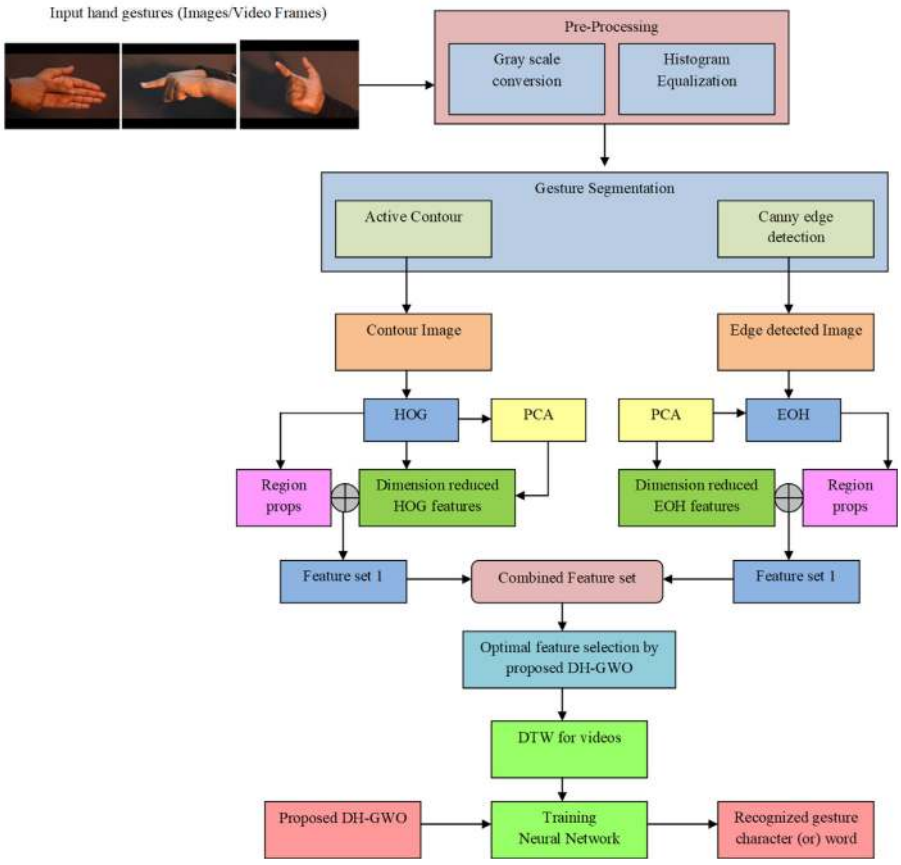

Fig. 1. Architectural model of proposed static and dynamic hand gesture recognition.

of dimension reduced HOG feature and region props features of contour image is represented as FZ_1 and the combination of dimension reduced EOH feature and region props features of the edge detected image is represented as FZ_2. Finally, the optimally selected features of FZ_1 and FZ_2 is denoted as $FZ^{optimal}$.

3.3. Image preprocessing

In the developed hand gesture recognition model, the image pre-processing is done by greyscale conversion and histogram equalization.

(a) **Grey scale conversion**: To form grey scale image, the colour intensity E^j of all pixels in colour image HE^l to all the trees in the random forest. At each node of the tree, the set of pixels are split for the stored binary test ϕ^*, and passed either to the left or right of the child node until reaching a leaf node. By sending through ith pixel in all trees in the forest, ith pixel ends in a set of leaf nodes LF_i. The maintained values in LF_i are the leaf node $\hat{I}^l$ and the covariance matrix $\sum^l$. With these values, the conversion of RGB colours of ith pixel to grey scale value GV^i, the mathematical equation is denoted in Eq. (1).

$$GV^i = \sum_{l \in LF_i} \omega^l dec\left(\hat{I}^l\right) \tag{1}$$

In Eq. (1), the normalized decimal value of $\hat{I}^l$ is denoted as $dec\left(\hat{I}^l\right)$, and ω^l denotes the confidence weight, defined by $\omega^l = \frac{1}{Trace\left(\sum^l\right)}$. Thus the grey scale image HE^{grey} is obtained.

(b) **Histogram equalization**: This technique [26] is used to change the intensity of an image for enhancing the image brightness. Consider HE^{grey} as the known image using k_l by k_m matrix of integer pixel intensities among 0 to 1, and PIV denotes the number of possible intensity values, regularly 256. Eq. (2) denotes the normalized histogram NH of E^j with a bin for each possible intensity.

$$NH = \frac{number\ of\ pixels\ with\ density\ he}{total\ number\ of\ pixels} \tag{2}$$

In Eq. (2), $he = 0, 1, \ldots, PIV - 1$. Eq. (3) denotes the histogram equalized image, which is denoted as E^j_{Hist}, in which the term $floor()$ rounds down to the closer integer.

$$E^j_{Hist} = floor((PIV - 1) \sum_{he=0}^{E^j_{grey_{(i,K)}}} NH) \tag{3}$$

Thus, the final pre-processed image by histogram equalization is denoted as E^j_{Hist} and this is used for the segmentation process.

4. Procedure for gesture segmentation and feature extraction

4.1. Gesture segmentation

In this section, segmentation is done using active contour models, and canny edge detection.

Active contour model: It [27] is determined as the energy minimizing spline that has the capacity of detecting different features in an image. The curve is a flexible curve, which is robustly altered in required objects or edges in the image, which consists of a set of control points that are associated with straight lines. In addition to this, whether the active contour is opened or closed curve is also checked. The active contour model is defined by Eq. (4).

$$\vec{w}(n) = \left(\vec{x}(n), \vec{y}(n)\right) \tag{4}$$

In the above equation, the term $x(n)$ and $y(n)$ indicates the co-ordinates of x and y, the control point's normalized index is denoted as n. Moreover, the energy function, which gives information regarding active contours that consists of external and internal energy. Internal energy makes the curve compact and constrained its acuminous deflections. Later, the external forces make the curve in the direction of the border of the object. The internal energy is denoted in Eq. (5), in which the adjustable constant is denoted as α that specifies the continuity, and the adjustable constant referring to contour curving is indicated by β. The elastic and bending energies are denoted in Eqs. (6) and (7).

$$ENG^{\text{int}} = ENG^{elas} + ENG^{bnd} = \alpha(n)\left|\frac{dw}{dn}\right|^2 + \beta(n)\left|\frac{d^2w}{dn^2}\right| \tag{5}$$

$$ENG^{elas} = \int_n \alpha\left(\vec{w}(n) - \vec{w}(n-1)\right)^2 dn \tag{6}$$

$$EN^{bnd} = \int_n \beta\left(\vec{w}(n-1) - \vec{w}(n) + \vec{a}(n+1)\right)^2 dn \tag{7}$$

Moreover, Eq. (8) indicates the minimized functional energy, in which the curve's internal energy is denoted as ENG^{int}, the picture energy is indicated by ENG^{img}, and the external limitations are denoted as ENG^{con}.

$$ENG^*_{snake} = \int_0^1 ENG^{snake}(w(n))\, dn = \int_0^1 \left\{ENG^{\text{int}}(w(n)) + ENG^{img}(w(n)) + ENG^{con}(w(n))\right\} dn \tag{8}$$

Thus, the image from active contour-based gesture segmentation is denoted as $HE^{contour}$.

Canny edge detection: This method [28] is a significant approach for extracting structural data, which noticeably decreases the amount of information to be processed. It is also called as optimal detector, which aims for fulfilling the three normal scenarios of edge detection. The variance among the real and the detected edge pixels have to be minimized. The detected edge must be marked only once not several times. The steps of canny edge detection algorithm are given below.

1. By using a filter, the noise present in the image is eliminated. Here, the Gaussian filter is employed for eliminating the noise. For example, the Gaussian kernel filter of size 5 is denoted in Eq. (9).

$$KR = \frac{1}{159}\begin{bmatrix} 2 & 4 & 5 & 4 & 2 \\ 4 & 9 & 12 & 9 & 4 \\ 5 & 12 & 5 & 12 & 5 \\ 4 & 9 & 12 & 9 & 4 \\ 2 & 4 & 5 & 4 & 2 \end{bmatrix} \tag{9}$$

2. A gradient image needs to be found. Gradients along x and y directions are computed by the convolution masks of 3 $\times$ 3 sizes is given in Eqs. (10), and (11).

$$Gr_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} \tag{10}$$

$$Gr_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix} \tag{11}$$

3. Moreover, the gradient strength and direction of edges are computed as per Eqs. (12), and (13). In this Gr_x and Gr_y are the gradients along with the directions of x and y.

$$Gr = \sqrt{Gr_x^2 + Gr_y^2} \tag{12}$$

$$\theta = \arctan\left(\frac{Gr_y}{Gr_x}\right) \tag{13}$$

4. The last step of this model uses two thresholds: those are called as upper and lower. When the pixel gradient value is more than the upper threshold, the respective pixel is taken as an edge pixel. When the pixel gradient value is less than the lower threshold, the pixel is rejected. When the pixel gradient value is between the upper and lower thresholds, the pixel will be accepted only when it is associated with the pixel, which is exceeding the upper threshold.

After the segmenting the edge of the hand gestures by the canny edge detection algorithm, the image is represented as HE^{edge}.

4.2. Feature extraction

The feature extraction is performed with the two segmented images. One is a contour image $HE^{contour}$, and the other is edge detected image HE^{edge}. Initially, the HOG feature is extracted from $HE^{contour}$ and EOH features are extracted from HE^{edge}.

HOG features: It [29] is the local object appearance and the shape of an image is determined using the distribution of density distribution of gradients. Moreover, the application of the HOG feature description is attained by splitting an image into small regions named a cell. Each cell in this descriptor compiles a histogram of gradient direction to the pixel present in the cell. Also, this approach consists of four steps to extract the object. Initially, the first step is used to compute the gradient values by implementing 1-D centred for acquiring the point of different derivative mask in the direction of vertical or horizontal as given in Eqs. (14) and (15), respectively.

$$K_y = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \tag{14}$$

$$K_x = \begin{bmatrix} -1 & 0 & 1 \end{bmatrix} \tag{15}$$

When the object image is denoted as Im, the terms x and y derivatives can be acquired by the convolution operation, which is denoted in Eq. (16). To compute the magnitude of the gradient and the corresponding equation is denoted in Eq. (17), and the gradient orientation is given by Eq. (18).

$$\begin{aligned} \mathrm{Im}_x &= \mathrm{Im}_x * K_x; \\ \mathrm{Im}_y &= \mathrm{Im}_y * K_y; \end{aligned} \tag{16}$$

$$|HG| = \sqrt{\mathrm{Im}_x^2 + \mathrm{Im}_y^2} \tag{17}$$

$$\theta = \arctan\frac{\mathrm{Im}_y}{\mathrm{Im}_x} \tag{18}$$

The other step is the spatial orientation binning, which has a function for acquiring the result of cell histogram by a voting procedure. In the final phase, block normalization is done by the L2 norm, which is denoted in Eq. (19).

$$bn = \frac{bn}{\sqrt{\|bn\|^2 + \varepsilon^2}} \tag{19}$$

Hence, FZ^{HOG} are the finally extracted HOG features from the contour image $HE^{contour}$.

EOH features: It [30] is a gradient-orientation based feature. EOH employs multiple 1-D features and each feature characterizes one orientation to a single block at a time. Therefore, the EOH feature is combined with a boosted cascade algorithm for effective weak learner selection. The computation of EOH features starts by performing edge detection in an image. The gradient magnitude $HG(x, y)$ and the gradient orientation $\theta(x, y)$ at the point (x, y) in an image Im is calculated by Eqs. (20) and (21), respectively.

$$HG(x, y) = \sqrt{HG_x(x, y)^2 + HG_y(x, y)^2} \tag{20}$$

$$\theta(x, y) = \arctan\left(\frac{HG_y(x, y)}{HG_x(x, y)}\right) \tag{21}$$

In the above equations, the term $HG_{im}(x, y)$ is the gradients at the point (x, y) that can be found using Sobel masks is denoted in Eqs. (22) and (23), respectively. The gradient orientation is evenly split into L bins and the gradient orientation histograms $EH_L(Bn_{im})$ in each bin l of block Bn_{im} are acquired using Eqs. (24) and (25), respectively.

$$HG_x(x, y) = Sobel_x * \mathrm{Im}(x, y) \tag{22}$$

$$HG_y(x, y) = Sobel_y * \mathrm{Im}(x, y) \tag{23}$$

$$EH_L(Bn_{im}) = \sum_{(x,y)\in Bn_{im}} \psi_L(x, y) \tag{24}$$

$$\psi_L(x, y) = \begin{cases} HG(x, y) & \theta(x, y) \in bn_l \\ 0 & Otherwise \end{cases} \tag{25}$$

A set of L EOH features of a single block is determined as the ratio of bin value of a single orientation to the sum of all the bin values as denoted in Eq. (26).

$$EOH^k_{Bn_{im}} = \frac{EH_l(Bn_{im}) + \varepsilon}{\sum_{j=1}^{L} EH_j(Bn_{im}) + \varepsilon} \tag{26}$$

Thus, the features FZ^{EOH} is the features extracted from the edge detected image HE^{edge}.

Dimension reduced HOG and EOH feature by PCA [31]: For decreasing the dimension of features, the dimensionality reduction approach named PCA is used once the feature extraction process is performed. The features FZ^{HOG} and FZ^{EOH} are considered for dimension reduction. PCA [31] is the dimension reduced feature that employed a basic arithmetical standard to convert many possible interrelated restrictions into an unimportant count of variables known as major constituents. Also, it is an arithmetic representation that produces extensive feature depictions. The main objective of this PCA is to reduce the dimension of huge data space, which is needed to represent the data creatively. This takes place when a robust association presents between defined constraints. The job of PCA is to have the ability to remove redundant information, prediction, feature extraction, and data compression. The mathematical representation of PCA is shown below.

1. Mean: Consider $N_1, N_2 \cdots N_n$ denotes the arbitrary parameters for a sample of size n. The arbitrary constraint is denoted in Eq. (27) that is the average of the dataset.

$$\overline{N} = \frac{1}{n}\sum_{i=1}^{n} N_i \tag{27}$$

2. Standard Deviation: The distance from the database denotes a specific point that has to be indicated for analysing standard deviation. The computation of the square of the distance from each point of the data to the average of the set is performed. The sum of the whole values is given in Eq. (28) and later divided by the whole numbers existing in the set.

$$SD = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(N_i - \overline{N}\right)^2} \tag{28}$$

3. Covariance: The configuration of this is approximately similar to the variance configuration. If the properties of N_i and R_i for $i = 1, 2, \ldots, n$ are evaluated, the sample variances of N and R is denoted in Eq. (29).

$$Cov\,(N, R) = \frac{\sum_{i=1}^{n} \left(N_i - \overline{N}\right)\left(R_i - \overline{R}\right)}{n} \tag{29}$$

4. Eigen values and Eigen vectors of a matrix: It is required to identify the eigen values and eigen vectors. If S is a $n \times n$ matrix, the term $N \neq \vec{0}$ is an eigenvector of S, where the scalar is indicated by λ and $N \neq \vec{0}$ according to Eqs. (30) and (31).

$$[S]\,[N] = \lambda X \tag{30}$$

$$\det\left([S\,| - \lambda Y = 0\,]\right) \tag{31}$$

The dimension reduced HOG features by PCA are represented as FZ_{PCA}^{HOG} and the dimension reduced EOH features by PCA is represented as FZ_{PCA}^{EOH}.

Region props: The region props features such as major and minor axis length, area, Euler number, equivalent diameter, bounding box, eccentricity, centroid, solidicity, extent, extrema, convex area, and orientation are considered in the proposed model.

(a) Major axis length: "The length (in pixels) of the major axis of the ellipse that has the same second-moments as the region".

(b) Minor axis length: "The length (in pixels) of the minor axis of the ellipse that has the same normalized second central moments as the region".

(c) Area: "The actual number of pixels in the region".

(d) Euler Number: "It is equal to the number of objects in the region minus the number of holes in those objects".

(e) Equivalent diameter: "The diameter of a circle with the same area as the region".

(f) Bounding Box: "It is the smallest rectangle containing the region".

(g) Eccentricity: "The eccentricity is the ratio of the distance between the foci of the ellipse and its major axis length. The value is between 0 and 1".

(h) Centroid: "The centre of mass of the region. Note that the first element of Centroid is the horizontal coordinate (or x-coordinate) of the centre of mass, and the second element is the vertical coordinate (or y-coordinate). All other elements of Centroid are in order of dimension".

(i) Solidicity: "The proportion of the pixels in the convex hull that are also in the region".

(j) Extent: "The proportion of the pixels in the bounding box that are also in the region".

(k) Extrema: "The extrema points in the region are defined by each row of the matrix contains the x- and y-coordinates of one of the points".

(l) Convex area: "It defines the number of pixels in a convex image".

(m) Orientation: "The angle (in degrees) between the x-axis and the major axis of the ellipse that has the same second-moments as the region".

Moreover, the region props features of the contour image and edge detected image is denoted as $FZ_{props}^{contour}$ and FZ_{props}^{edge}, respectively. Hence, FZ_1 is the feature set 1 with the combination of dimension reduced HOG features and region props features of contour image as shown in Eq. (32). Similarly, FZ_2 is considered as the feature set with the combination of dimension reduced EOH features and region props features of the edge detected image as mentioned in Eq. (33).

$$FZ_1 = FZ_{PCA}^{HOG} + FZ_{props}^{contour} \tag{32}$$

$$FZ_2 = FZ_{PCA}^{EOH} + FZ_{props}^{edge} \tag{33}$$

Finally, FZ^* attains the combined feature set based ob Eq. (34) with the summation of FZ_1 and FZ_2.

$$FZ^* = FZ_1 + FZ_2 \tag{34}$$

From the collection of feature set FZ^*, the optimal features $FZ^{optimal}$ has to be selected with the help of a hybrid DH-GWO algorithm.

5. Optimal feature selection and detection for static and dynamic gesture recognition

5.1. Optimal feature selection

The optimal feature selection is done by the proposed DH-GWO. The entire features FZ^* is taken by the proposed DH-GWO, and the optimal features $FZ^{optimal}$ are generated.

5.2. Dynamic time warping

In the dynamic image (video), the training and testing should involve a few numbers of frames that are giving different information. Hence, the frames with repeated information are removed, which have less difference from the previous frame. The combined feature set should be considered for all frames taken for a single video. The DTW [18] model computes the disparity among two data series obtained at distinct times. A matrix consisting of the Euclidean distances at aligned points over two series is employed for measuring the minimal cost among the two series. Moreover, the selection of the direction to the shortest path is linked with certain rules and regulations. In specific, the movement is limited to vertical, horizontal, and diagonal directions. A weight is linked to each of these directions and the shortest path has to be inferior for a threshold to the two series that are required to be assumed as equivalent.

5.3. Optimized neural network

The optimally selected features $FZ^{optimal}$ is subjected to the optimized NN [32] for hand gesture recognition using both static as well as dynamic data. The structure of the NN consists of input, output, and hidden layers. The hidden layer's outcome is required for measuring the final outcome of the network.

$$\overline{G}^{(G)} = af\left(\widetilde{S}_{(\hat{T}q)}^{(G)} + \sum_{p=1}^{In(op)} \widetilde{S}_{(pq)}^{(G)} FR_i^*\right) \tag{35}$$

In Eq. (35), the activation function is denoted as af, and the number of input neurons is indicated by $In\,(op)$. The bias weight to the hidden neuron is denoted as $\widetilde{S}_{(\hat{T}q)}^{(G)}$ and the weight from the input neuron to the hidden neuron is denoted as $\widetilde{S}_{(pq)}^{(G)}$. The input and the hidden neurons are denoted as p and q, respectively. The input features are denoted as FR_i^*. The result of the output layer is expressed in Eq. (36).

$$\hat{H}_r = af\left(\widetilde{S}_{(\hat{T}r)}^{(H)} + \sum_{q=1}^{hdn} \widetilde{S}_{(qr)}^{(H)} \overline{G}^{(G)}\right) \tag{36}$$

In the above equation, the count of hidden neurons hdn and the output neurons are denoted as r. The bias weight from the output neuron is denoted as $\widetilde{S}_{(\hat{T}r)}^{(H)}$ and the weight from the hidden and the output neurons are indicated by $\widetilde{S}_{(qr)}^{(H)}$. The weight function $S_{weu}^{NN} = \left\{\widetilde{S}_{(\hat{T}q)}^{(G)}, \widetilde{S}_{(\hat{T}r)}^{(H)}, \widetilde{S}_{(pq)}^{(G)}, \widetilde{S}_{(qr)}^{(H)}\right\}$ is chosen optimally for producing

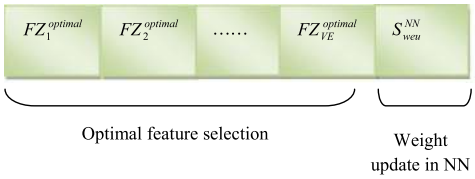

Fig. 2. Solution encoding for feature selection and classification.

better training in NN with minimum error variance as shown in Eq. (37).

$$FR2 = \arg\underset{\left\{S_{weu}^{NN}\right\}}{Min} \sum_{r=1}^{H(op)} \left|H_r - \hat{H}_r\right| \tag{37}$$

In Eq. (3), the actual and the predicted outcome is denoted as H_r and $\hat{H}_r$, respectively.

6. Developed Deer hunting-based Grey Wolf optimization for optimal gesture recognition

6.1. Objective model

The main objective model of the proposed hand gesture recognition is to maximize the recognition accuracy. The term accuracy is the "ratio of the observation of exactly predicted to the whole observations". The formula for accuracy is shown in Eq. (38), where TR^P, TR^N denotes the true positives, and true negatives, respectively. Moreover, FA^P and FA^N are false positives and false negatives, respectively.

$$Acc = \frac{TRP + TRN}{TRP + TRN + FAP + FAN} \tag{38}$$

The optimal feature selection and optimal trained NN for recognition are performed by the proposed DH-GWO. Hence, the objective model is defined in Eq. (39)

$$Ob = \underset{\left\{FZ_1^{optimal}, FZ_2^{optimal}, \ldots, FZ_{VE}^{optimal}, S_{weu}^{NN}\right\}}{\arg Max} (Acc) \tag{39}$$

6.2. Solution encoding

The optimal feature selection and the optimal trained classification takes the features and weight function as input that is to be optimized by the proposed DH-GWO. The solution encoding of the proposed optimization is shown in Fig. 2.

In Fig. 2, *VE* indicates the total number of features. The minimum and maximum bounding limit of feature selection are 0 and 1, in which 0 means the features are not selected, and 1 means they are selected.

6.3. Conventional DHOA

DHOA [33] is inspired to chase the deer by the hunting behaviour of the human. Conventional DHOA aims to obtain the optimum position to shoot the deer. Moreover, conventional DHOA has some features that are quite complex to humans for hunting the deer. The visual power of the deer is five times more than human beings. The hunter's population is initialized by *po*. The mathematical equation is denoted in Eq. (40). In this, the term *N* indicates the total count of hunters. The wind angle and the position angle of the deer are initialized to determine the best locations of the hunters. Consider circle as the search space, the wind angle follows the circle's circumference and the corresponding equation is denoted in Eqs. (41) and (42) indicates the position angle.

$$po = \{po_1, po_2, \ldots, po_N\} \quad 1 < a < N \tag{40}$$

$$\theta_{tis=} 2\Pi rd \tag{41}$$

$$\phi_{tis} = \theta + \Pi \tag{42}$$

In Eqs. (41) and (42), the term *rd* denotes the random number. The leader and the successor position are considered and they are expressed as po^{lr} and po^{sr}, respectively. The best solution of the hunter is chosen by the leader's position, and the next best solution is selected by the successor's position. Later, every hunter strives for being in the best location, and the process of update starts. Consequently, the encircling behaviour is given in Eq. (43).

$$po_{tis+1} = po^{lr} - A \cdot b \cdot \left|B \times po^{lr} - po^{tis}\right| \tag{43}$$

In the above equation, the coefficient vectors are indicated by *A* and *B*. Moreover, the random number is denoted as *b* based on wind angle and the values ranging from [0,2]. The coefficient vectors *A* and *B* are shown in Eqs. (44) and (45). Here, the parameter *d* ranging from [−1, 1], and in Eq. (45), the random number is represented as *c* and it lies in between 0 and 1. In update rule, the position angle is considered and the search space improvement is performed. Eq. (46) refers to the angle of visualization of prey. The update process related to position angle is denoted in Eq. (47) that is calculated based on the difference between the visual and the position angle.

$$A = \frac{1}{4} \log\left(tis + \frac{1}{tis_{\max}}\right) d \tag{44}$$

$$B = 2 \cdot c \tag{45}$$

$$vsa_{tis} = \frac{\Pi}{8} \times rd \tag{46}$$

$$df_{tis} = \theta_{tis} - vsa_{tis} \tag{47}$$

By using Eq. (48), the update process of position angle in the next iteration is done. Moreover, the position update is performed by considering the position angle as shown in Eq. (50).

$$\phi_{tis+1} = \phi_{tis} + df_{tis} \tag{48}$$

$$po_{tis+1} = po^{lr} - b \cdot \left|\cos(e) \times po^{lr} - po_{tis}\right| \tag{49}$$

During the exploration phase, the behaviour is taken by regulating the vector *B*. Thus, the position update is performed based on the position of the successor instead of the first best solution. The global search mathematical equation is denoted in Eq. (50).

$$po_{tis+1} = po^{sr} - A \cdot b \cdot \left|B \times po^{sr} - po_{tis}\right| \tag{50}$$

In each iteration, the position update is performed until the best solution is obtained.

6.4. Conventional GWO

The classical GWO [34] is inspired by the hunting behaviour of wolves. These are also called as apex predators. They are mostly observed in sets including 5–12 in each cluster following leadership mechanism. The hierarchy of each set of wolves is divided into four classes like alpha α, beta β, delta δ, and omega ω. The prey is encircled by the grey during hunting. This kind of behaviour is mathematically represented in Eqs. (51) and (52).

$$EP = \left|C \cdot po_{py}(tis) - po(tis)\right| \tag{51}$$

$$po(tis + 1) = po_{py}(tis) - D \cdot EP \tag{52}$$

In the above equations, the position vector of the prey is denoted as po_{py}, and the coefficient vectors are expressed as *C*

and D. The coefficient vectors are computed by Eqs. (53) and (54). Here, the random numbers are denoted as rad_1 and rad_2, these are ranging from 0 to 1. The components of g is decreased from 2 to 0 in the series of iteration.

$$C = 2 \cdot rad_2 \tag{53}$$

$$D = 2g \cdot rad_1 - g \tag{54}$$

The hunting is majorly controlled by alpha, next to the hunting done by beta and delta. The hunting behaviour of wolves for mathematical simulation is considered as alpha, beta, and delta, which have complete data related to the prospective positions of the prey. Thus, the first three best solutions attained are needed for saving and creating the other search agents to update the positions for the best search agent position. The mathematical equation to hunt is denoted in Eqs. (55), (56), and (57).

$$\begin{aligned} EP_\alpha &= |C_1 \cdot po_\alpha - po|, \\ EP_\beta &= \left|C_2 \cdot po_\beta - po\right|, \\ EP_\delta &= |C_3 \cdot po_\delta - po| \end{aligned} \tag{55}$$

$$\begin{aligned} po_1 &= po_\alpha - D_1 \cdot (EP_\alpha), \\ po_2 &= po_\beta - D_2 \cdot \left(EP_\beta\right), \\ po_3 &= po_\delta - D_3 \cdot (EP_\delta) \end{aligned} \tag{56}$$

$$po(tis+1) = \frac{po_1 + po_2 + po_3}{3} \tag{57}$$

6.5. Proposed DH-GWO

The inspiration of traditional GWO is the hunting behaviours of the wolves. Moreover, these wolves have the capability of recognizing the location of the wolf in a better way and encircling the prey. The advantages of existing GWO are its simple implementation, and it requires only some amount of storage space. However, there are few drawbacks such as bad local searching capability, less solving accuracy, and low convergence rate. The fast convergence rate of the DHOA algorithm tries to make the GWO best by hybridizing this concept. The major advantage of traditional DHOA is its capability to solve the problem with more strength and has many ways to find a global solution. Despite the conventional GWO, the solutions po_1, po_2 and po_3 is updated by the DHOA concept. Here, po_1 is updated by Eq. (43), po_2 is updated by Eq. (50), and po_3 is updated by Eq. (49) of DHOA. The pseudo code of the proposed DH-GWO is shown in Algorithm 1.

Algorithm 1: Pseudocode of Proposed DH-GWO Algorithm

```
Grey wolf population is initialized by po_Pop(Pop = 1, 2, ..., n)
g, C, D are initialized
Fitness of each search agent need to be computed
po_α be the best search agent
po_β be the second best search agent
po_δ be the third best search agent
while (tis < Max_iterations)
   for each search agent
      Update po_1 using Eq. (43) of DHOA
      Update po_2 using Eq. (50) of DHOA
      Update po_3 using Eq. (49) of DHOA
      Update current search agent position by Eq. (55)
   end for
   Update g, C, and D
   Compute fitness of all search agents
   Update po_α, po_β, and po_δ
   tis = tis + 1
end while
return po_α
```

7. Results and discussions

7.1. Experimental setup

The proposed hand gesture recognition system using static and dynamic data was implemented on Python, and the performance analysis was carried out. The parameters have been tuned by the trial and error method, which is a fundamental method of problem solving. It is characterized by repeated, varied attempts, which are continued until success. Here, the benchmark ISL dataset [35,36] was considered, which consists of both static and dynamic images. For performing the experiment, the population size was fixed as 10 and the maximum number of iterations was taken as 25. The performance of the proposed DH-GWO-NN model was compared over conventional models such as PSO-NN [37], GWO-NN [34], WOA-NN [38], and DHOA-NN [33]. Also, the performance of the proposed DH-GWO-NN was compared over the existing machine learning algorithms like KNN [39], SVM [40], NB [41], and NN [32]. The measures such as accuracy, sensitivity, specificity, precision, FPR, FNR, NPV, FDR, F1 score, and MCC was considered for analysing the performance.

7.2. Performance measures

The description of the performance measures considered for analysing the performance is given below.

(a) Accuracy: It is described in Eq. (38).

(b) Sensitivity: It measures "the number of true positives, which are recognized exactly".

$$Sen = \frac{TRP}{TRP + FAN} \tag{58}$$

(c) Specificity: It measures "the number of true negatives, which are determined precisely".

$$Spe = \frac{TRN}{FAP} \tag{59}$$

(d) Precision: It is "the ratio of positive observations that are predicted exactly to the total number of observations that are positively predicted".

$$\Pr e = \frac{TRP}{TRP + FAP} \tag{60}$$

(e) FPR: It is computed as "the ratio of the count of false positive predictions to the entire count of negative predictions".

$$FPR = \frac{FAP}{FAP + TRN} \tag{61}$$

(f) FNR: It is "the proportion of positives which yield negative test outcomes with the test".

$$FNR = \frac{FAN}{TRN + TRP} \tag{62}$$

(g) NPV: It is the "probability that subjects with a negative screening test truly do not have the disease".

$$NPV = \frac{FAN}{FAN + TRN} \tag{63}$$

(h) FDR: It is "the number of false positives in all of the rejected hypotheses".

$$FDR = \frac{FAP}{FAP + TRP} \tag{64}$$

(i) F1 score: It is defined as the "harmonic mean between precision and recall. It is used as a statistical measure to rate performance".

$$F1score = \frac{Sen \bullet \Pr e}{\Pr e + Sen} \tag{65}$$

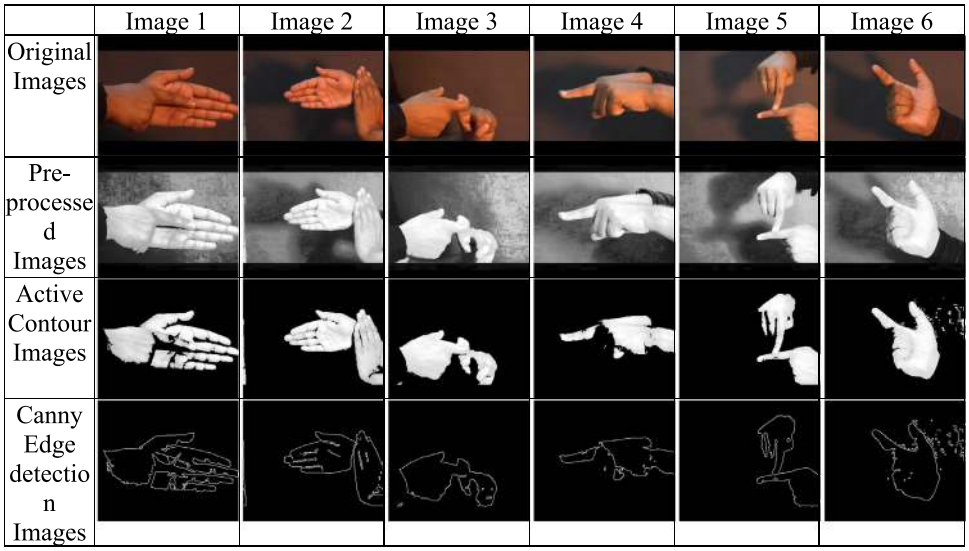

Fig. 3. Experimental results of pre-processing and segmentation for hand gesture recognition.

Table 2
Overall performance analysis of proposed and conventional heuristic-based NN models for hand gesture in static basis.

Algorithms	Accuracy	Sensitivity	Specificity	Precision	FPR	FNR	NPV	FDR	F1 score	MCC
PSO-NN [37]	0.968182	0.71	0.994	0.922078	0.006	0.29	0.994	0.077922	0.80226	0.793212
GWO-NN [34]	0.966364	0.69	0.994	0.92	0.006	0.31	0.994	0.08	0.788571	0.780124
WOA-NN [38]	0.967273	0.7	0.994	0.921053	0.006	0.3	0.994	0.078947	0.795455	0.786688
DHOA-NN [33]	0.964545	0.67	0.994	0.917808	0.006	0.33	0.994	0.082192	0.774566	0.766869
DH-GWO-NN	0.97	0.71	0.996	0.946667	0.004	0.29	0.996	0.053333	0.811429	0.805216

Table 3
Overall performance analysis of proposed and conventional heuristic-based NN models for hand gesture in dynamic basis.

Algorithms	Accuracy	Sensitivity	Specificity	Precision	FPR	FNR	NPV	FDR	F1 score	MCC
PSO-NN [37]	0.874286	0.142857	0.955556	0.263158	0.044444	0.857143	0.955556	0.736842	0.185185	0.130302
GWO-NN [34]	0.862857	0.171429	0.939683	0.24	0.060317	0.828571	0.939683	0.76	0.2	0.12943
WOA-NN [38]	0.865714	0.228571	0.936508	0.285714	0.063492	0.771429	0.936508	0.714286	0.253968	0.182547
DHOA-NN [33]	0.851429	0.2	0.92381	0.225806	0.07619	0.8	0.92381	0.774194	0.212121	0.130728
DH-GWO-NN	0.897143	0.228571	0.971429	0.470588	0.028571	0.771429	0.971429	0.529412	0.307692	0.279108

(j) MCC: It is a "correlation coefficient computed by four values".

$$MCC = \frac{TRP \times TRN - FAP \times FAN}{\sqrt{(TRP + FAP)\,(TRP + FAN)\,(TRN + FAP)\,(TRN + FAN)}} \tag{66}$$

7.3. Segmentation analysis

The experimental results of the hand gesture recognition are shown in Fig. 3, which involves pre-processed images, active contour images, and canny edge detection images.

7.4. Effect of optimized NN

The performance analysis of the proposed and existing heuristic-based NN for learning percentage using images is shown in Fig. 4. In Fig. 4(a), the accuracy of the improved DH-GWO-NN is precisely determined for all the learning percentages. At learning percentage 35%, the accuracy of the developed DH-GWO-NN is 0.4% better than PSO-NN, 0.5% better than DHOA-NN, and 0.6% better than WOA-NN. Moreover, the precision of the suggested DH-GWO-NN is performing well when considering any of the learning percentages and it is shown in Fig. 4(d). The precision of the recommended DH-GWO-NN is 1.4% advanced than PSO-NN, 1.5% advanced than DHOA-NN, and 3% advanced than WOA-NN when considering the learning percentage as 35. Also, the FDR of the introduced DH-GWO-NN is superior when compared to all the other methods at all learning percentages. When considering the learning percentage as 65, the FDR of the presented DH-GWO-NN is 16.6% superior to DHOA-NN and it is shown in Fig. 4(h). From Fig. 4(i), the F1 score of the improved DH-GWO-NN is 0.9% enhanced than WOA-NN, 1.5% enhanced than PSO-NN, 4.1% advanced than GWO-NN, and 6.1% enhanced than DHOA-NN at learning percentage 85. The performance analysis of the developed DH-GWO-NN and the conventional models using videos concerning learning percentages is shown in Fig. 5. In Fig. 5(a), the accuracy of the suggested DH-GWO-NN is determined accurately for all the learning percentages. It is 2.1% upgraded than GWO, 2.6% upgraded than WOA, and 4.3% upgraded than DHOA-NN at learning percentage 35. Moreover, the precision of the suggested DH-GWO-NN is 18.6% surpassed than GWO-NN, 27.1% surpassed than WOA-NN, and 32.8% surpassed than DHOA-NN when considering the learning percentage as 35 and it is shown in Fig. 5(d). In Fig. 5(j), the MCC of the recommended DH-GWO-NN is 50% and 100% better than WOA, and DHOA-based NN, respectively. Therefore, the results of the improved DH-GWO-NN are found to be superior to conventional models and it is well suitable for recognizing both dynamic and static hand gestures.

In Table 2, the overall performance of the developed DH-GWO-NN and conventional methods for the images are depicted. From Table 2, the accuracy of the suggested DH-GWO-NN is predicted

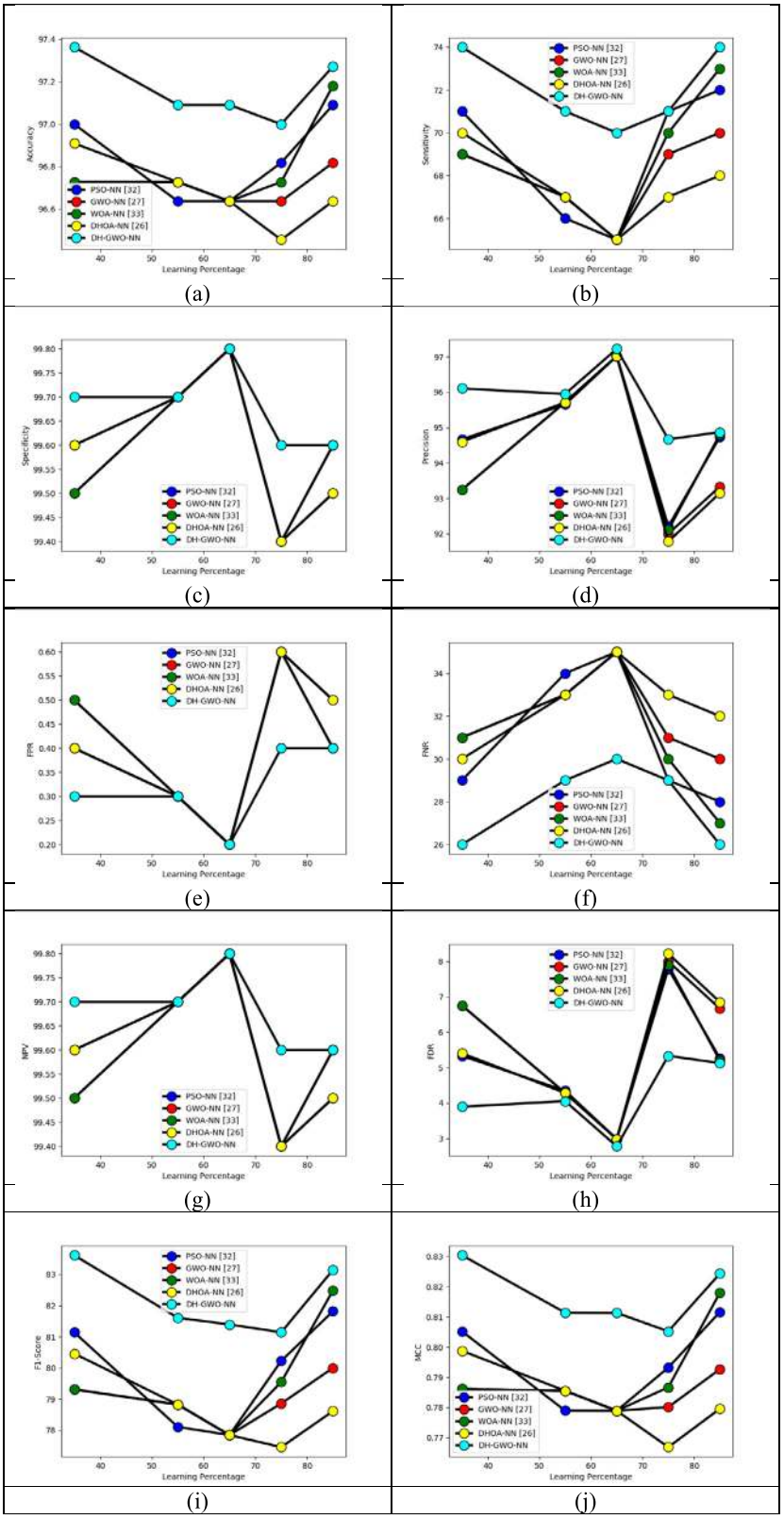

Fig. 4. Performance analysis of the proposed and conventional heuristic-based NN for hand gesture recognition using static images by varying the learning percentages for the measures (a) accuracy, (b) sensitivity, (c) specificity, (d) precision, (e) FPR, (f) FNR, (g) NPV, (h) FDR, (i) F1 score, and (j) MCC.

accurately when compared over other methods. The accuracy of the improved DH-GWO-NN is 0.1% better than PSO-NN, 0.3% better than GWO-NN, 0.2% better than WOA-NN, and 0.5% better than DHOA. Similarly, the precision of the developed DH-GWO-NN is accurately defined when compared with other models. It is 2.6% superior to PSO-NN, 2.8% superior to GWO-NN, 2.7% superior to WOA-NN, and 3.1% superior to DHOA-NN. Similarly, Table 3

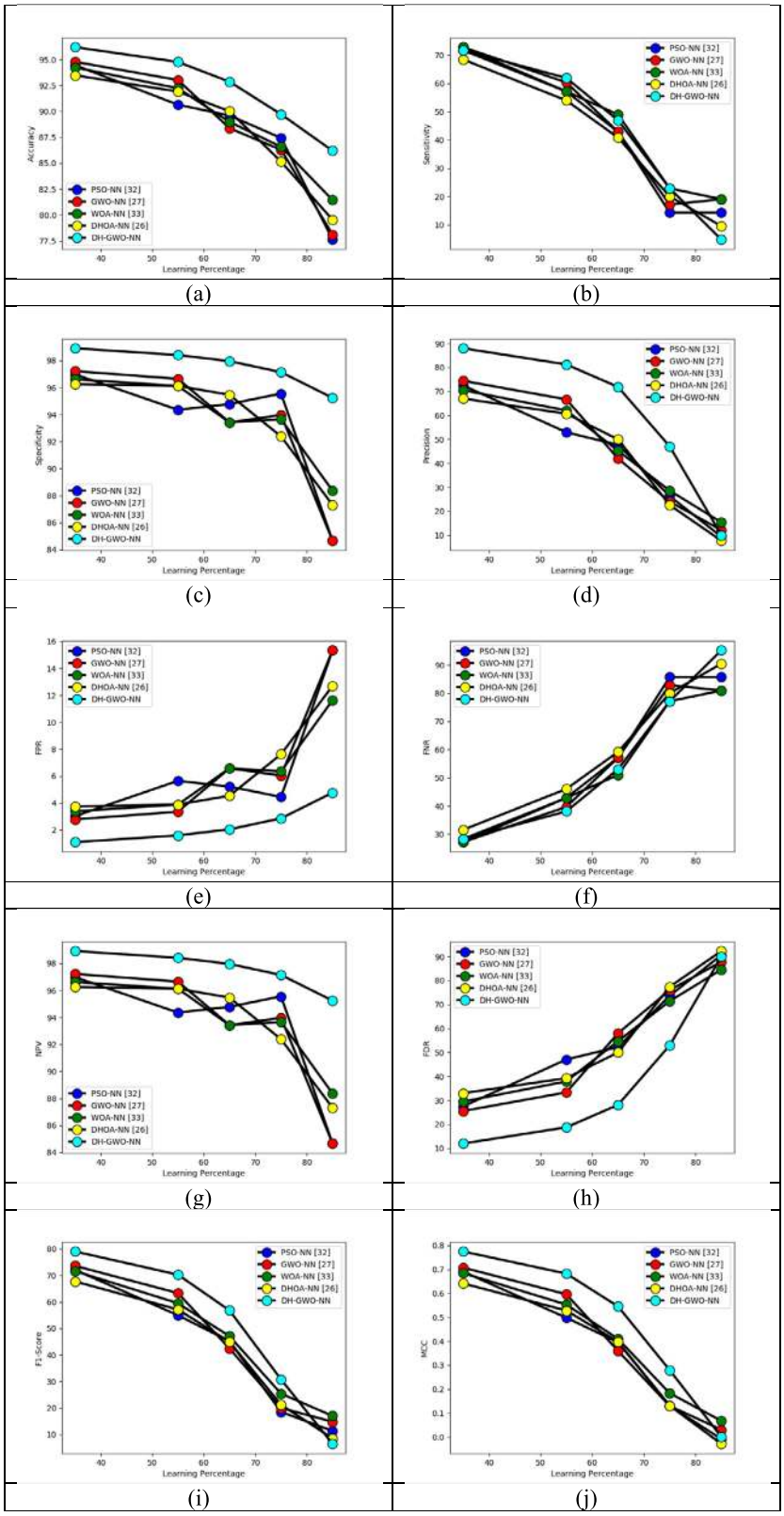

Fig. 5. Performance analysis of the proposed and conventional heuristic-based NN for hand gesture recognition using dynamic images by varying the learning percentages for the measures (a) accuracy, (b) sensitivity, (c) specificity, (d) precision, (e) FPR, (f) FNR, (g) NPV, (h) FDR, (i) F1 score, and (j) MCC.

describes the overall performance of the suggested DH-GWO-NN and the existing techniques for videos. In Table 3, the accuracy of the introduced DH-GWO-NN is 2.6% advanced than PSO-NN, 3.9% advanced than GWO-NN, 3.6% advanced than WOA-NN, and 5.3% advanced than DHOA-NN. Moreover, the precision of the implemented DH-GWO-NN is 78.8% improved than PSO-NN, 96%

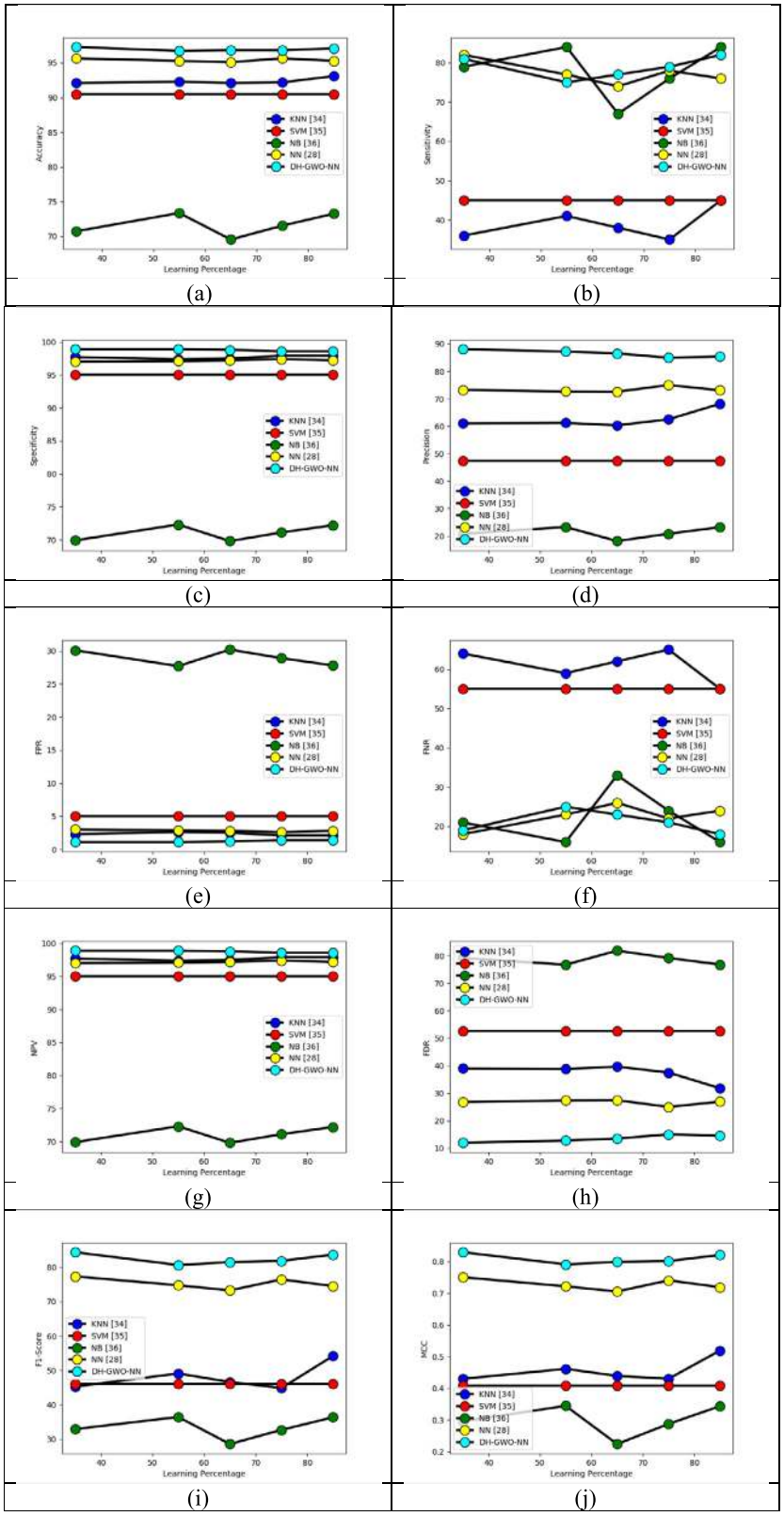

Fig. 6. Performance analysis of the proposed and conventional machine learning algorithms for hand gesture recognition using static images by varying the learning percentages for the measures (a) accuracy, (b) sensitivity, (c) specificity, (d) precision, (e) FPR, (f) FNR, (g) NPV, (h) FDR, (i) F1 score, and (j) MCC.

improved than GWO-NN, 64.7% improved than WOA-NN, and 52% improved than DHOA-NN. Thus, it is concluded that the proposed DH-GWO-NN is performing well in hand gesture recognition in static as well as dynamic basis.

7.5. Performance analysis over conventional machine learning

The analysis of the proposed DH-GWO-NN and the conventional classifiers using images for hand gesture recognition with

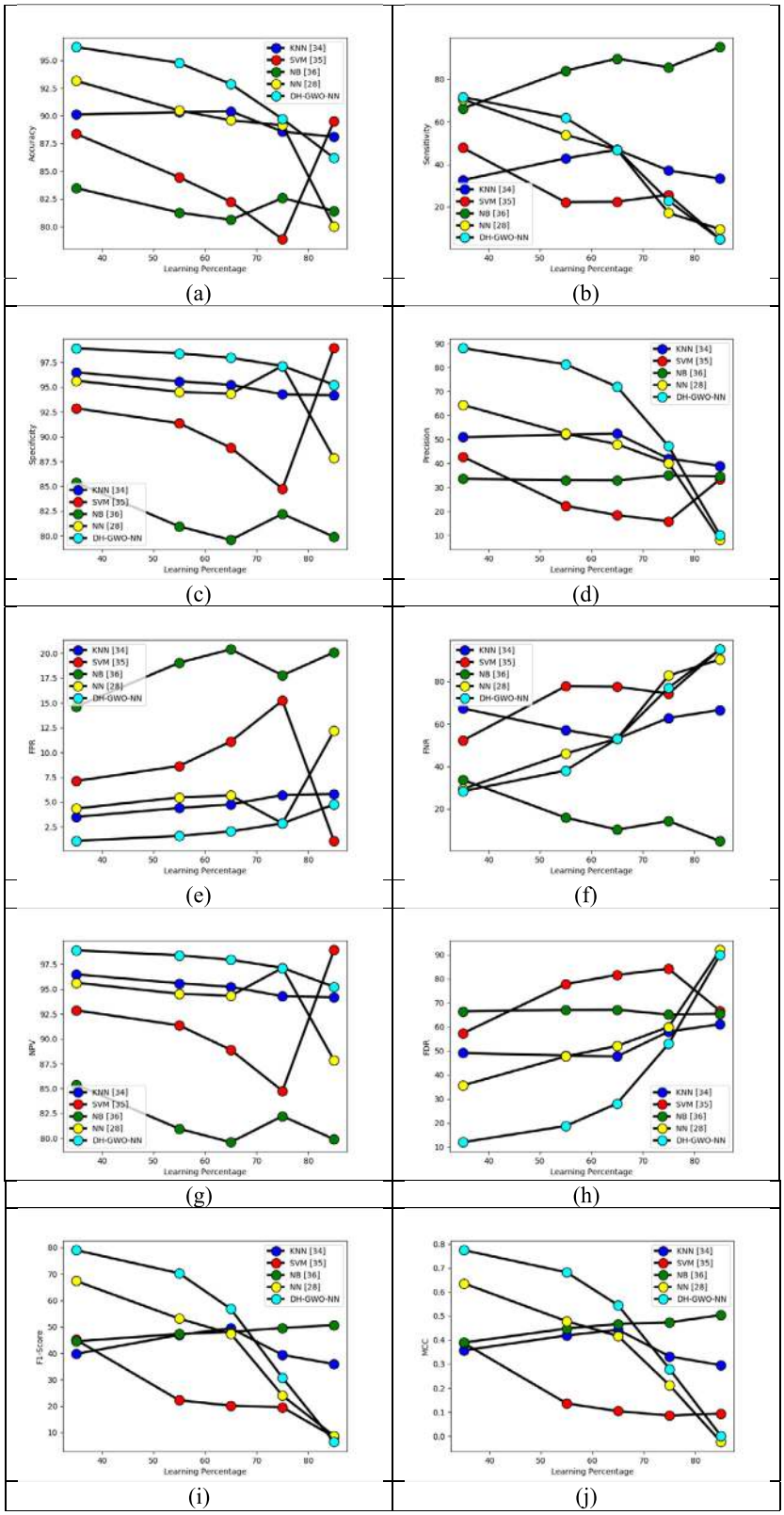

Fig. 7. Performance analysis of the proposed and conventional machine learning algorithms for hand gesture recognition using dynamic images by varying the learning percentages for the measures (a) accuracy, (b) sensitivity, (c) specificity, (d) precision, (e) FPR, (f) FNR, (g) NPV, (h) FDR, (i) F1 score, and (j) MCC.

respect to learning percentage is given in Fig. 6. In Fig. 6(a), the accuracy of the improved DH-GWO-NN is defined exactly for all the learning percentages. From Fig. 6(a), the accuracy of the implemented DH-GWO-NN at learning percentage 35 is 2% enhanced than NN, 5.9% enhanced than KNN, 8.8% SVM, and 38% enhanced than NB. Moreover, the precision of the suggested DH-GWO-NN is correctly determined for all the learning percentages,

Table 4
Overall performance analysis of proposed and conventional machine learning algorithms for hand gesture in static basis.

Algorithms	Accuracy	Sensitivity	Specificity	Precision	FPR	FNR	NPV	FDR	F1 score	MCC
KNN [39]	0.921818	0.35	0.979	0.625	0.021	0.65	0.979	0.375	0.448718	0.43028
SVM [40]	0.945455	0.4	1	1	0	0.6	1	0	0.571429	0.614295
NB [41]	0.715455	0.76	0.711	0.208219	0.289	0.24	0.711	0.791781	0.326882	0.287562
NN [32]	0.964545	0.67	0.994	0.917808	0.006	0.33	0.994	0.082192	0.774566	0.766869
DH-GWO-NN	0.97	0.71	0.996	0.946667	0.004	0.29	0.996	0.053333	0.811429	0.805216

Table 5
Overall performance analysis of proposed and conventional machine learning algorithms for hand gesture in dynamic basis.

Algorithms	Accuracy	Sensitivity	Specificity	Precision	FPR	FNR	NPV	FDR	F1 score	MCC
KNN [39]	0.885714	0.371429	0.942857	0.419355	0.057143	0.628571	0.942857	0.580645	0.393939	0.331847
SVM [40]	0.788571	0.257143	0.847619	0.157895	0.152381	0.742857	0.847619	0.842105	0.195652	0.085118
NB [41]	0.825714	0.857143	0.822222	0.348837	0.177778	0.142857	0.822222	0.651163	0.495868	0.473414
NN [32]	0.891429	0.171429	0.971429	0.4	0.028571	0.828571	0.971429	0.6	0.24	0.211604
DH-GWO-NN	0.897143	0.228571	0.971429	0.470588	0.028571	0.771429	0.971429	0.529412	0.307692	0.279108

which is shown in Fig. 6(d). The precision of the proposed DH-GWO-NN is 16.4% better than NN, 39.3% better than KNN, 77% better than SVM, and 71.7% better than NB at learning percentage 55. At learning percentage 35, the FPR of the implemented DH-GWO-NN is 50% superior to KNN, 66.6% superior to NN, 80% superior to SVM, and 96.6% superior to NB, which is shown in Fig. 6(e). Moreover, the classification performance analysis of the developed DH-GWO-NN and the machine learning algorithms using videos concerning learning percentage is given in Fig. 7. The accuracy of the introduced DH-GWO-NN is 4.3% improved than NN, 7.7% improved than KNN, 10.2% improved than SVM, and 16.8% improved than NB at learning percentage 35, and it is shown in Fig. 7(a). Also, from Fig. 7(d), the precision of the proposed DH-GWO-NN defined the true observations from all the observations correctly for all the learning percentages. At learning percentage 55, the precision of the improved DH-GWO-NN is 60.7% advanced than NN, 60.9% advanced than NB, and 73.1% advanced than SVM. Hence, it is proved that the developed DH-GWO-NN is performing well in recognizing the hand gestures accurately.

In Table 4, the overall classification analysis of the proposed DH-GWO-NN and the existing classifiers using images is tabulated. The accuracy of the modified DH-GWO-NN is defined accurately. It is 4.9% better than KNN, 2.5% better than SVM, 35.5% better than NB, and 0.5% better than NN. Moreover, the precision of the introduced DH-GWO-NN is 51.4% improved than KNN, 5.3% improved than SVM, 78% improved than NB, and 3.1% improved than NN. Likewise, the overall classification performance of developed DH-GWO-NN and traditional classifiers using videos is tabulated in Table 5. From Table 5, the accuracy of the improved DH-GWO-NN is 1.2% upgraded than KNN, 13.7% upgraded than SVM, 8.6% upgraded than NB, and 0.6% upgraded than NN. Similarly, the precision of the suggested DH-GWO-NN is 12.2% surpassed than KNN, 66.4% surpassed than SVM, 34.9% surpassed than NB, and 17.6% surpassed than NN. Therefore, it is confirmed that the proposed DH-GWO-NN is well suitable for hand gesture recognition.

8. Conclusion

The proposed model has developed an effective hand gesture recognition models considering both static and dynamic datasets for ISL. In static type, the images were considered for processing, and the video frames were employed for processing the dynamic type. At first, the grey scale conversion and histogram equalization was performed in the pre-processing phase. Later, the image was segmented by an active contour model and canny edge detection. Both the contour image and the edge detected image was deployed in the feature extraction phase, where HOG, EOG features were extracted from the contour image and the edge detected images, respectively. At last, these features were added, and the optimal feature selection was done for selecting the unique feature providing various information with less correlation. Moreover, the recognition classifier named NN was used, in which the training model was employed for updating the network weight. DTW approach helped for eliminating the redundant frames present in the video, and for decreasing the time used for testing. From the experimental results, the accuracy of the developed DH-GWO-NN was 4.9% better than KNN, 2.5% better than SVM, 35.5% better than NB, and 0.5% better than NN. Moreover, the precision of the introduced DH-GWO-NN is 51.4% improved than KNN, 5.3% improved than SVM, 78% improved than NB, and 3.1% improved than NN. Finally, it is concluded that the proposed DH-GWO is performing well for hand gesture recognition using both static and dynamic types.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] R. Xie, J. Cao, Accelerometer-based hand gesture recognition by neural network and similarity matching, IEEE Sens. J. 16 (11) (2016) 4537–4545.
[2] S. Poularakis, I. Katsavounidis, Low-complexity hand gesture recognition system for continuous streams of digits and letters, IEEE Trans. Cybern. 46 (9) (2016) 2094–2108.
[3] S.Y. Kim, H.G. Han, J.W. Kim, S. Lee, T.W. Kim, A hand gesture recognition sensor using reflected impulses, IEEE Sens. J. 17 (10) (2017) 2975–2976.
[4] Z. Zhang, Z. Tian, M. Zhou, Latern: Dynamic continuous hand gesture recognition using FMCW radar sensor, IEEE Sens. J. 18 (8) (2018) 3278–3289.
[5] J.P. Sahoo, S. Ari, D.K. Ghosh, Hand gesture recognition using DWT and F-ratio based feature descriptor, IET Image Process. 12 (10) (2018) 1780–1787.
[6] Y. Liu, Y. Zhang, M. Zeng, Novel algorithm for hand gesture recognition utilizing a wrist-worn inertial sensor, IEEE Sens. J. 18 (24) (2018) 10085–10095.
[7] Joyeeta Singha, Amarjit Roy, Rabul Hussain Laskar, Dynamic hand gesture recognition using vision-based approach for human–computer interaction, Neural Comput. Appl. 29 (4) (2018) 1129–1141.
[8] Dong-Luong Dinh, Sungyoung Lee, Tae-Seong Kim, Hand number gesture recognition using recognized hand parts in depth images, Multimedia Tools Appl. 75 (2) (2016) 1333–1348.
[9] Meng Xing, Jing Hu, Zhiyong Feng, Yong Su, Weilong Peng, Jinqing Zheng, Dynamic hand gesture recognition using motion pattern and shape descriptors, Multimedia Tools Appl. 78 (8) (2019) 10649–10672.
[10] Hong Cheng, Zhongjun Dai, Zicheng Liu, Yang Zhao, An image-to-class dynamic time warping approach for both 3D static and trajectory hand gesture recognition, Pattern Recognit. 55 (2016) 137–147.

[11] G. Li, S. Zhang, F. Fioranelli, H. Griffiths, Effect of sparsity-aware time–frequency analysis on dynamic hand gesture classification with radar micro-doppler signatures, IET Radar, Sonar Navig. 12 (8) (2018) 815–820.
[12] Anna K. Lekova, D. Ryan, Reggie Davidrajuh, Fingers and gesture recognition with kinect v2 sensor, Inform. Technol. Control 14 (3) (2016).
[13] A.R. Varkonyi-Koczy, B. Tusor, Human–computer interaction for smart environment applications using fuzzy hand posture and gesture models, IEEE Trans. Instrum. Meas. 60 (5) (2011) 1505–1514.
[14] Q. Chen, N.D. Georganas, E.M. Petriu, Hand gesture recognition using haar-like features and a stochastic context-free grammar, IEEE Trans. Instrum. Meas. 57 (8) (2008) 1562–1571.
[15] Chen-Chiung Hsieh, Dung-Hua Liou, Novel haar features for real-time hand gesture recognition using SVM, J. Real-Time Image Process. 10 (2) (2015) 357–370.
[16] K. Kollorz, J. Penne, J. Hornegger, A. Barke, Gesture recognition with a time-of-flight camera, Int. J. Intell. Syst. Technol. Appl. 5 (3) (2008) 334–343.
[17] K. Rimkus, A. Bukis, A. Lipnickas, S. Sinkevičius, 3D human hand motion recognition system, in: 2013 6th International Conference on Human System Interactions (HSI), Sopot, 2013, pp. 180–183.
[18] G. Plouffe, A. Cretu, Static and dynamic hand gesture recognition in depth data using dynamic time warping, IEEE Trans. Instrum. Meas. 65 (2) (2016) 305–316.
[19] Di Wu, Lionel Pigou, Pieter-Jan Kindermans, Nam Do-Hoang Le, Ling Shao, Joni Dambre, Jean-Marc Odobez, Deep dynamic neural networks for multimodal gesture segmentation and recognition, IEEE Trans. Pattern Anal. Mach. Intell. 38 (8) (2016) 1583–1597.
[20] Oyebade K. Oyedotun, Adnan Khashman, Deep learning in vision-based static hand gesture recognition, Neural Comput. Appl. 28 (12) (2017) 3941–3951.
[21] Zhongxu Hu, Youmin Hu, Jie Liu, Bo Wu, Dongmin Han, Thomas Kurfess, 3D separable convolutional neural network for dynamic hand gesture recognition, Neurocomputing 318 (2018) 151–161.
[22] Jingren Tang, Hong Cheng, Yang Zhao, Hongliang Guo, Structured dynamic time warping for continuous hand trajectory gesture recognition, Pattern Recognit. 80 (2018) 21–31.
[23] D. Lee, W. You, Recognition of complex static hand gestures by using the wristband-based contour features, IET Image Process. 12 (1) (2018) 80–87.
[24] G. Li, R. Zhang, M. Ritchie, H. Griffiths, Sparsity-driven micro-doppler feature extraction for dynamic hand gesture recognition, IEEE Trans. Aerosp. Electron. Syst. 54 (2) (2018) 655–665.
[25] Hao Tang, Hong Liu, Wei Xiao, Nicu Sebe, Fast and robust dynamic hand gesture recognition via key frames extraction and feature fusion, Neurocomputing 331 (2019) 424–433.
[26] R. Dorothy, R.M. Joany, Joseph Rathish, S. Santhana Prabha, Susai Rajendran, Joseph, Image enhancement by histogram equalization, Int. J. Nano Corros. Sci. Eng. 2 (2015) 21–30.
[27] Marián Bakoš, Active contours and their utilization at image segmentation, in: 5th Slovakian-Hungarian Joint Symposium on Applied Machine Intelligence and Informatics, 2007.
[28] R. Pradeep Kumar Reddy, Chiluka Nagaraju, I. Raja Sekhar Reddy, Canny scale edge detection, 2016.
[29] Bambang Sugiarto, Esa Prakasa, Riyo Wardoyo, Ratih Damayanti, Listya Krisdianto, Mustika Dewi, Hilman F. Pardede, Yan Rianto, Wood identification based on histogram of oriented gradient (HOG) feature and support vector machine (SVM) classifier, in: 2017 2nd International Conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE), 2017.
[30] Yingdong Ma, Xiankai Chen, Liu Jin, George Chen, A monocular human detection system based on EOH and oriented LBP features, in: International Symposium on Visual Computing, 2011, pp. 551–562.
[31] Xingfu Zhang, Xiangmin Ren, Two dimensional principal component analysis based independent component analysis for face recognition, in: 2011 International Conference on Multimedia Technology, Hangzhou, 2011, pp. 934-936.
[32] F. Fernández-Navarro, M. Carbonero-Ruz, D. Becerra Alonso, M. Torres-Jiménez, Global sensitivity estimates for neural network classifiers, IEEE Trans. Neural Netw. Learn. Syst. 28 (11) (2017) 2592–2604.
[33] G. Brammya, S. Praveena, N.S. Ninu Preetha, R. Ramya, B.R. Rajakumar, D. Binu, Deer hunting optimization algorithm: A new nature-inspired meta-heuristic paradigm, 2019.
[34] Seyedali Mirjalili, Seyed Mohammad Mirjalili, Andrew Lewis, Grey wolf optimizer, Adv. Eng. Softw. 69 (2014) 46–61.
[35] A. Nandy, S. Mondal, J.S. Prasad, P. Chakraborty, G.C. Nandi, Recognizing & interpreting Indian sign language gesture for human robot interaction, in: The proceeding of ICCCT'10, IEEE Xplore Digital Library, 2010, pp. 712-717.
[36] A. Nandy, S. Mondal, J.S. Prasad, P. Chakraborty, G.C. Nandi, Recognition of isolated Indian sign language gesture in real time, in: The Book of (Information Processing and Management) Springer LNCS-CCIS, Vol. 70, 2010, pp. 102–107.
[37] M.E.H. Pedersen, A.J. Chipperfield, Simplifying particle swarm optimization, Appl. Soft Comput. 10 (2) (2010) 618–628.
[38] Seyedali Mirjalili, Andrew Lewis, The whale optimization algorithm, Adv. Eng. Softw. 95 (2016) 51–67.
[39] Yewang Chen, Xiaoliang Hu, Wentao Fan, Lianlian Shen, Zheng Zhang, Xin Liu, Jixiang Du, Haibo Li, Yi Chen, Hailin Li, Fast density peak clustering for large scale data based on kNN, Knowl.-Based Syst. (2019) Available online.
[40] Shuang Yu, Kok KiongTan, Ban LeongSng, Shengjin, Alex Tiong HengSia, Lumbar ultrasound image feature extraction and classification with support vector machine, Ultrasound Med. Biol. 41 (10) (2015) 2677–2689.
[41] A. Sanchis, A. Juan, E. Vidal, A word-based Naïve Bayes classifier for confidence estimation in speech recognition, IEEE Trans. Audio Speech Lang. Process. 20 (2) (2012) 565–574.