IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 23 May 2023, accepted 21 June 2023, date of publication 26 June 2023, date of current version 15 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3289389

RESEARCH ARTICLE

Smart Healthcare Hand Gesture Recognition Using CNN-Based Detector and Deep Belief Network

MOHAMMED ALONAZI[1], HIRA ANSAR[2], NAIF AL MUDAWI[3], SAUD S. ALOTAIBI[4], NOUF ABDULLAH ALMUJALLY[5], ABDULWAHAB ALAZEB[3], AHMAD JALAL[2], JAEKWANG KIM[6], (Member, IEEE), AND MOOHONG MIN[7]

[1]Department of Information Systems, College of Computer Engineering and Sciences, Prince Sattam bin Abdulaziz University, Al-Kharj 16273, Saudi Arabia
[2]Faculty of Computer Science and AI, Air University, Islamabad 44000, Pakistan
[3]Department of Computer Science, College of Computer Science and Information System, Najran University, Najran 55461, Saudi Arabia
[4]Information Systems Department, Umm Al-Qura University, Makkah 24382, Saudi Arabia
[5]Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh 11671, Saudi Arabia
[6]Social Innovation Convergence Program, University College, Sungkyunkwan University, Seoul 03063, Republic of Korea
[7]School of Convergence/Convergence Program for Social Innovation, Sungkyunkwan University, Seoul 03063, Republic of Korea

Corresponding author: Moohong Min (iceo@skku.edu)

This research is supported and funded by Sungkyunkwan University. This research is supported and funded by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R410), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. This study is supported via funding from Prince Satam bin Abdulaziz University project number (PSAU/2023/R/1444). The authors are thankful to the Deanship of Scientific Research at Najran University for funding this work under the Research Group Funding program grant code (NU/RG/SERC/12/6).

ABSTRACT Gesture recognition in dynamic images is challenging in computer vision, automation and medical field. Hand gesture tracking and recognition between both human and computer must have symmetry in real world. With advances in sensor technology, numerous researchers have recently proposed RGB gesture recognition techniques. In our research paper, we introduce a reliable hand gesture tracking and recognition model that is accurate despite any complex environment, it can track and recognise RGB dynamic gestures. Firstly, videos are converted into frames. After images light intensity adjustment and noise removal, images are passed through CNN for hand gesture extraction. Then from the extracted hand, features are extracted from full hand. Neural gas and locomotion thermal mapping are extracted to make the feature vector. The feature vector are then passed through the fuzzy optimiser to reduce the uncertainties and the fuzziness. The optimised features are then passed to the classifier Deep Belief Network (DBW) for the classification of the gestures. Egogesture and Jester datasets are used for the validation of proposed systems. The experimental results over Egogesture and Jester datasets demonstrate overall accuracies of 90.73% and 89.33% respectively. The experiments proves our system readability and suitability of our proposed model with the other state of the arts model.

INDEX TERMS Convolution neural network, neural gas, thermal locomotion mapping, fuzzy logic, deep belief network, hand detection and tracking.

I. INTRODUCTION

In our daily routine, human and computer interaction is now an essential part which helps us in multiple ways. With the advancement in the computer technology and hardware, Human Computer Interaction (HCI) serves with its best techniques [1]. Hand gesture recognition is one of the most important and on-demand area for researchers now, as it provides a simple way to interact with computers. The interaction between human and computer is intuitive and user friendly. Humans used joysticks, remote controls and special devices for interaction but users should be trained for such devices. Hand gesture interface are more user friendly and flexible systems for human computer interaction. Humans

The associate editor coordinating the review of this manuscript and approving it for publication was Giovanni Pau.

and computer interaction can be helpful in different domains such as home appliances, robotics, medical, virtual environment, and in communication [2], [3], [4], [5]. The detection of hand gesture is a crucial point to cater in research area. As in different circumstances, hand gesture recognition rate varies. Complex background and luminosity affects its recognition rate. In field like medical errors lead to loss of life, researchers are figuring error-free error free recognition in every situation. For such problems a lot of different models have been proposed which works efficiently, but to improvise all issues regarding recognition is a difficult task to achieve [6].

In hand gesture recognition some generic steps are used such as data acquisition, hand detection, feature extraction and then the hand is classified. The data is collected using sensors and cameras, also depth sensors, such devices collect static and dynamic gestures [7], [8], [9]. Many researcher have proposed techniques for detecting gestures including machine learning and deep learning methods. Now in research, people have tried to overcome the limitations. In our research work we have proposed a solution to improve the finding in research. The proposed work efficiently in medical and home automation area [10]. It also serves as a communication tool for humans and computers.

The main contributions of the paper are:

- Hand gestures are recognised with deep learning algorithm such as deep belief network to get better accuracy, and to detect on large scale with complex background.
- Features of the hand such as neural gas and locomotive thermal mapping are novel features used to identify gestures for classification.
- Features fuzziness is removed for making the feature vectors to extract the gesture with accuracy.
- A comparative study is performed to check the system reliability and to compare our model with the other state of the art algorithms.

In Section II Related work is given. Section III includes the methodology to get the accurate results. In Section IV Experimentation has done, to validate our system. Finally, Section V presents the conclusion and future work.

II. RELATED WORK

Researchers have conducted various studies to implement hand gesture recognition in medical field to assist the medical staff. This section includes the recent methods and techniques to understand the importance of gestures in medical field. The litrature review is divided into two main data acquisition techniques i.e. hand gesture detection using RGB sensors and via marker sensors.

A. MEDICAL GESTURES ACQUISITION VIA MARKER SENSORS

In [11] inspected patient gesture using series learning through remote monitoring. Intermediate and structured features are extracted in their model for dependable gesture recognition (DGR). The proposed system provided 94.92% accuracy with reduction time 4.97 s and 4.93s for DGR. Bargellesi et al. [12] presented a model which used wearable motion capture sensor to recognize hand gesture via random forest. The hand movements are analyzed for features extraction at different time intervals. They have used different gestures dataset to perform experimental evaluation. Cho et al. [13] proposed gesture recognition in operating room using a personalized classifier which improved gesture recognition. They took 5 gestures namely: hover, grab, click, one peak, and two peaks and extracted thirty features for classification, those features were passed to support vector machine (SVM) and Naïve Bayes (NB) classifiers to recognize computer-aided surgical procedure. There model gave 99.58%±0.06, and 98.74%±3.64 accuracy using SVM and NB respectively. In [14] Tavakoli et al. improvised a wearable sensor with double surface EMG. The used 4 gestures for classification via 2 EMGs channels on flexors and extensors. The gestures were fed to SVM and 95% accuracy was obtained with less fault tolerance. In [15] Zhang et al. acquired hand gesture key points using a convolutional pose machine (CPM) and for hand gesture recognition they used Fuzzy Gaussian mixture models (FGMM). The FGMM also eliminated the non-gesture classes and used vital points for classifying gestures. In [16] Kopuklu et al. proposed a system to diagnose online dynamic hand gesture which enhanced gesture analysis efficiency. For gesture classification they used two-level hierarchy structure via CNN. There proposed method was workable on gestures with no start and complete point, the reorganized gesture can only be identified at a single time, and gesture recognition equipment should be under budget and space provided.

B. MEDICAL GESTURES ACQUISITION VIA VIDEO SENSORS

In [17] Hamed et al. presented recovery room monitering system in which videos were converted into frames. The frames helped in extracting color seperation based approach for feature selection. The features were trained through hidden markov model to classify the gesture. The system succesfully attained 91.36% accuracy. In [18] Van den Berg et al. proposed a method which detected 6 gestures such as: one open-hand,punching, pointing up, L-shaped motion, pointing to the camera, and two open-handed collected via cameras. They succesfully detected complex 3D gestures from background and recongnized gestures. In [19] Ameur et al. recognise 11 hand gestures using various classifiers: SVM, nearest neighbor (NN), decision tree (DT), random forest (RF), AdaBoost, linear discriminant analysis and the multi-layer perceptron. The best classification results were achieved via SVM and multi-layer perceptron attaining the accuracy of 91.73% and 89.91% respectively. In [20] Ameur et al. evaluated there work on RIT dataset extracting spatial, frequential and spatio-frequential features. Those features were collected from the fingertips and the palm center. The feature vector was then fed into SVM for classification. They

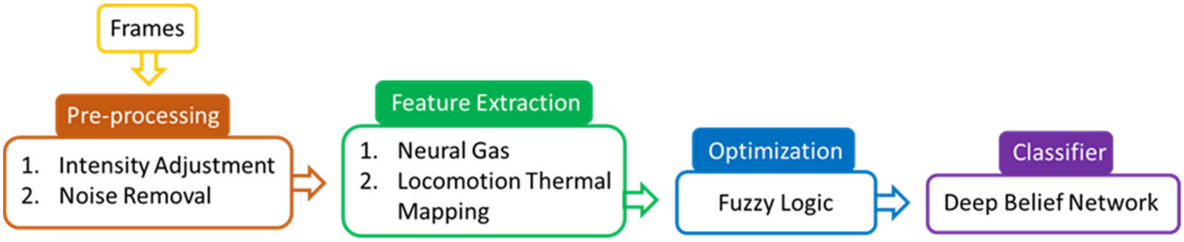


FIGURE 1. The system architecture of the proposed.

achieved 93% accuracy over their own dataset. Li et al. [21] proposed a model for rehabilitation after stroke. 17 people performed 7 different gestures named as: extension-flexion of wrist, close and spread fingers, fingertip tap, radial-ulnar deviation, fingertip touch, palm rotation and finger mass extension-flexion. These gestures were evaluated using k-fold cross validation method. The SVM and KNN provided 97.29% and 97.71% respectively. In [22] Dutta et al. proposed a method for gesture operated system which localized hand by bounding box and class labels was assigned. They detected keypoints on hand for support drawing and path traversal. They produced 89.6% accuracy for OUHANDS dataset. There system proved the gestures were affective for rehabilitation.

III. PROPOSED SYSTEM METHODOLOGY

In this paper, we proposed a novel approach for hand gesture recognition that recognizes and label them in videos. Initially, videos are taken and images frames are extracted. These frames are passed through CNN which successfully detects the hand gesture with accuracy. From the detected hand, features are computed. For features neural gas and locomotion thermal mapping are used. After that the extracted features are optimized with fuzzy logic and given to classifier DBN for recognition. Fig. 1 presents overview of the proposed architecture.

A. IMAGE PRE-PROCESSING

Images contains unwanted information due to many reasons. These unwanted pixels to be removed to detect the object accurately. The images are pre-processed to remove unwanted pixels for the adjustment of luminosity and noise removal. First the image intensity is adjusted of the original images to improve contrast [23]. Gamma correction is used to make image brighter. It provides the best result with improving image quality. Gama correction possesses non-linear behavior and its power law definition is as follows:

$$G_x = WL_y^{\gamma} \tag{1}$$

where L_y is the non-negative input image which is raised to γ and is multiplied with the constant value W, $W = 1.G_x$ is the output received after gamma correction lies between range 0 and 1.

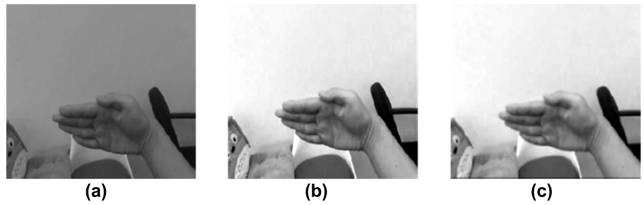


FIGURE 2. Some examples from Egogesture dataset: (a) original image; (b) intensity adjustment and (c) denoised image.

After gamma correction image is passed through non-local mean filter which process image to denoise. Non-local mean filter reduce loss of image quality by taking mean value of all pixels and the weight is calculated to verify the similarity between original and target pixel [24]. Γ is the area of the original image, a and b are the two points in the image, then the formulae will be:

$$\mu(a) = \frac{1}{D(a)} \int_{\Gamma} \omega(b) f(a, b)\, db \tag{2}$$

where $\mu(a)$ is the filtered value at point a, $\omega(b)$ presents the unfiltered value at point b, $f(a, b)$ is the weighted function which determine the closeness found in the both points a and b. Figure 2(c) shows the resultant denoised image and its histogram [25]. Figure 3 presents the histogram representation of those resultant image.

B. HAND DETECTION USING CNN

Hand detection involves different methods such as skin detection, thermal, deep learning methods. Hand segmentation need accurate extraction of gesture for successful recognition. In this research, we considered CNN for hand detection as it provides promising results in any background without error. In CNN, three convolution layers, three max-pooling layer, one fully connected layer and output layer is present [26]. Figure 4 presents the architecture of proposed CNN model for hand detection. The convolution layer CV_1 is at start which is consist of input matrix [27], [28], [29], [30]. The convolution layer have 32 kernels, size of 1 * 13. The convolution matrix is calculated using formulae

$$CV_i^{j-1}(m, n) = \sum_{y=1}^{x} \Omega\left(m\left(n - y + \frac{x+1}{2}\right)\right) U_i^j(y) + \alpha_i^j \tag{3}$$

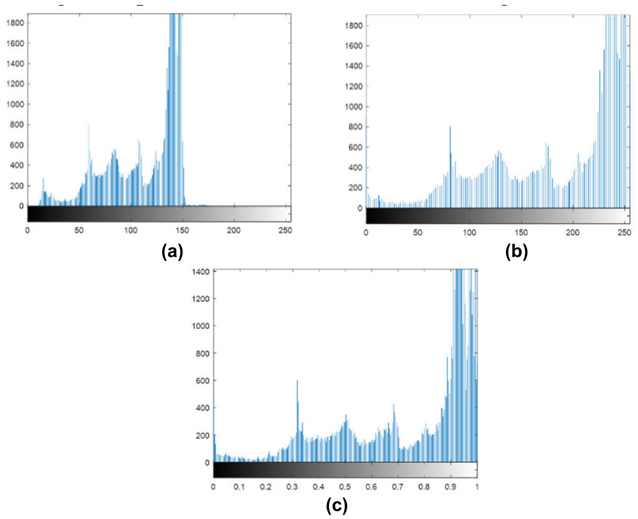


FIGURE 3. **Histogram representation: (a) original image; (b) intensity adjustment and (c) denoised image.**

where (m, n) are coordinates for convolution layer, Ω presents last layer map and kernel size. U_i^j presents the j^{th} kernel for layer i^{th}. α_i^j presents the bias of the convolution layer. The resultant output $CV_i^{j-1}(m, n)$ is passed through the first layer of max-pooling layer M_1. The sum of the bian and weight between convolution layer and max pool layer is calculated using ReLU function and passed to the next layer. This layer down sample the results using sliding window of 1 * 2 size. The pooling layer result can be formulated using the equation:

$$M_i^{j-1}(m, n) = \max\left(CV_i^j\left(m\left((n-1) \times (r, s)\right)\right)\right) \tag{4}$$

where $(r-1)$ is the output layer, s is the kernel and (m, n) are coordinates. $1 \leq r \leq s$ and i is the size of pooling window. The first pooling layer results are then passed through the second convolution layer having 64 kernels and the whole procedure is again repeated but with 128 kernels. At last fully connected layer is obtained as

$$\begin{aligned} Ful_u^{j+1} = \sum\nolimits_{y=1}^{x} \Omega\left(m\left(n - y + \frac{x+1}{2}\right)\right) U_i^j(y) + \alpha_i^j \\ + \alpha_i^j\left(\sum_t d_t^j U_{ju}^j + \alpha_u^j\right) \end{aligned} \tag{5}$$

where $d_t^j U_{ju}^j$ presents the matrix with weights. The convergence plot of CNN for egogesture dataset is given with 250 epochs. Also the detected hand is presented in Fig 5.

C. FEATURE EXTRACTION USING NEURAL GAS

Neural gas is self-organizing neural map, it is capable of determining the neighborhood data space with the help of ranking neighborhood vectors [31], [32], [33]. It is composed of neurons i has a weight vector presented as $N(v)$ which forms clusters. The position of each neuron shows abrupt movement during training. Every single neuron is associated with feature vector randomly. From the feature vector random data v is selected which forms a neural gas network. From all the weight vectors, the Euclidean distance from that random data vector v is computed. With the calculated distance the centers are adjusted using that selected data vector. Every feature vector adapts itself by the following computation

$$N_{f_i}^{T+1} = N_{f_i}^T + \varepsilon . e^{-i/\lambda} . \left(v - N_{f_i}^T\right), i = 0, \ldots, W-1 \tag{6}$$

where probability distribution $N(v)$ of data vector v with finite number of sets $s_f, f = 1, \ldots.., W$. A data vector v from probability distribution $N(v)$ is presented from each time step T. The distance order is determined from feature vector of the given data v. Suppose i_o be the index of the closed feature vector, i_1 be the second and i_{w-1} be the distant to the data vector v. ε presents adaptation step size and λ presents neighborhood range. After most of the adaptation step the data space is covered with feature vector with minimum errors [34]. Fig. 6 presents the structure of neural gas over egogesture dataset gesture.

D. FEATURE EXTRACTION USING LOCOMOTION THERMAL MAPPING

In locomotion thermal mapping, thermal maps are captured for hand movements of different gestures from one frame to another [35]. The hand movements speed differs comparing two altered frames, which implies that the gestures having the greater speed will be given higher heat values. The hand with greater speed are shown in yellowish color. And the areas of hand which were not involved in any movement while performing the gesture are presented in green color. The range is set between 0 to 8000 to show th eheat values of geatures in thermal mapping. The set of values involved in greater movement are extracted for classification. The thermal mapping is represented by the following equation:

$$LTM(x) = \sum\nolimits_0^j \ln S(j) \tag{7}$$

where the extracted values are stored in 1D vector x, j presents index matrix values and S presents the values extracted from j. Fig. 7 presents the result of locomotion thermal mapping on different gestures.

E. FUZZY LOGIC OPTIMISATION

Extracted features are passed through optimiser for feature selection and reduction. Fuzzy Optimisation (FO) solves the problem of uncertainty in parameters using membership function [36]. FO allows the partial truth and partial false mathematical representation. The parameters are not represented as part of set or not but the degree of membership is allocated to each data set. Membership function defines the range allocation between the ranges 0 to 1. For the removal of fuzziness, Defuzzification method acquires output with non-fuzzy values. The formulae is given as:

$$\Delta Fuzz = \frac{\sum_{i=1}^{q} \alpha_i(q) \times o_i}{\sum_{i=1}^{p} \alpha_i(q)} \tag{8}$$

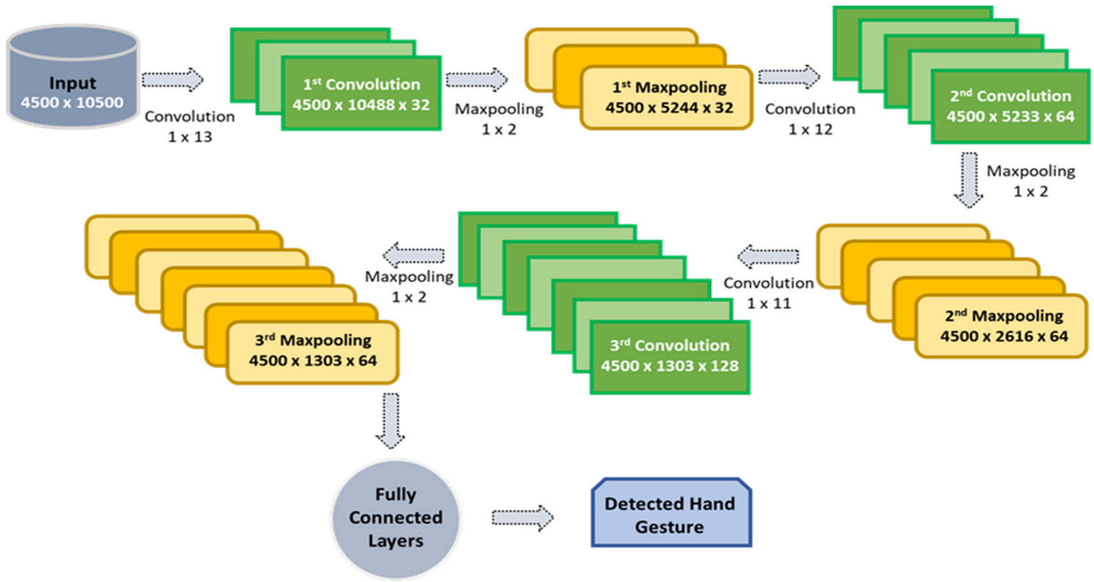
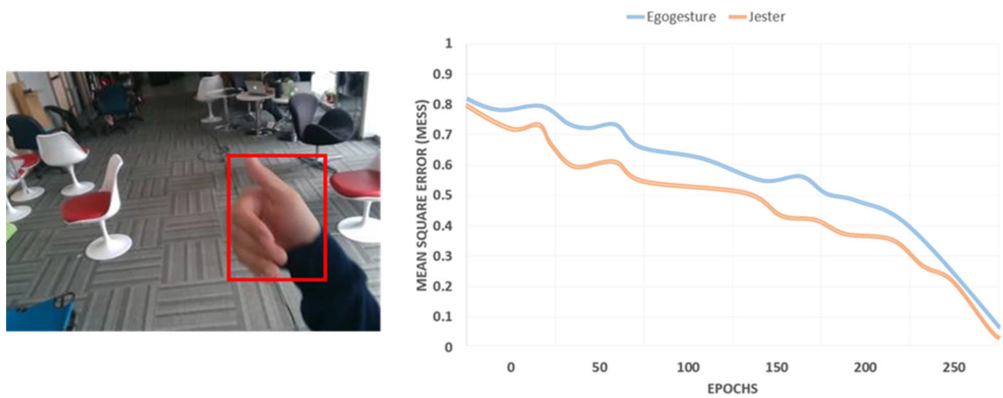
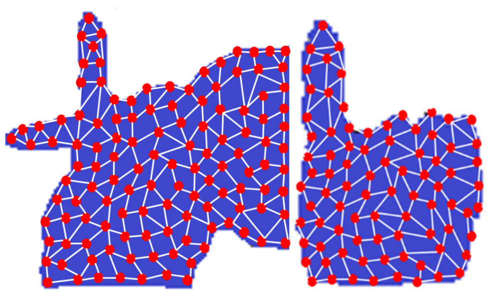


FIGURE 4. The architecture of proposed CNN model for hand detection.



(a) (b)

FIGURE 5. Hand detection using CNN: (a) detected hand and (b) convergence plot over egogesture and jester dataset.



(a) (b)

FIGURE 6. Neural gas over egogesture dataset gesture classes: (a) cross index finger and (b) click with index finger.

where $\alpha_i(q)$ presents accomplishment in ith rule, o_i represents the height rule. The surface function can express the relation between fuzzy inputs and outputs.

(a) (b) (c)

FIGURE 7. LTM over egogesture dataset gesture.

F. DEEP BELIEF NETWORK

After feature extraction the DBN is used over two datasets i.e. Egogesture and Jester for gesture recognition. The DBN, a type of classifier with multiple number of restricted Boltzmann machines (RBMs), which serves as the basic unit and also BP neural network (BPNN) [37]. RBMs reduces the feature dimensionality, extract and filter the features before feeding it to the BPNN. RBMs includes two layers i.e. hidden

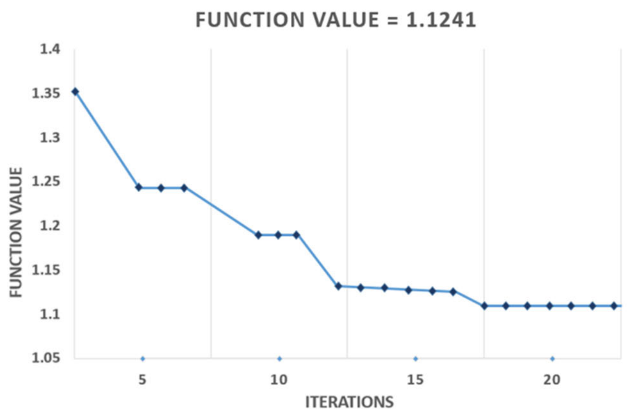


FIGURE 8. Fuzzy optimization result over egogesture datasets.

layer and visible layer. The particular layers have no connection within the RBM units and no data interaction is seen [38]. Fig. 9 presents the graphical representation of the RBMs and Figure 10 represents the architecture of DBN. The energy function $E(x, y)$ of combined units, the hidden unit (y) and the trained sample (x) through probability is defined as:

$$E(x, y) = \sum_{k=1}^{l} a_k x_k - \sum_{h=1}^{m} b_h y_h - \sum_{h=1}^{m} \sum_{k=1}^{l} x_k c_{kh} y_h \tag{9}$$

where a and b represents biases value of the hidden and visible layer and c represents the weights matrix assigned to visible and hidden units for connection [39]. When the input unit is determined the hidden layer probability is computed as:

$$PH(y = 1|x, \theta) = \prod_{k=0}^{l} \frac{1}{(1 + \exp(-C_k^T x - d_k))} \tag{10}$$

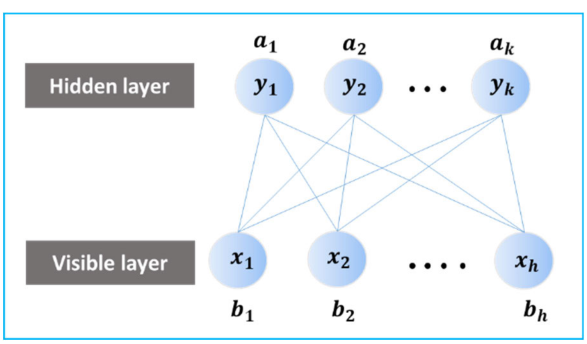


FIGURE 9. The graphical the graphical representation of the RBMs.

After the hidden unit is determined the visible unit starts being activate, the probability computed of visibility unit is:

$$PV(x = 1|y, \theta) = \prod_{h=0}^{l} \frac{1}{(1 + \exp(-C_h^T y - e_h))} \tag{11}$$

The joint probability of both units x and y are:

$$p(x, y) = \frac{1}{FG(\theta)} r^{-E(x,y)} \tag{12}$$

$$FG(\theta) = \sum_{x,y} r^{-E(x,y)} \tag{13}$$

where $FG(\theta)$ presents the regulation constant. The weights between both units are updated based on positive gradient subtracted from negative gradient $x'y'^W$ times alongwith learmimg rate (β). The weight updation is computed as

$$\Delta W = \beta xy^W - x'y'^W \tag{14}$$

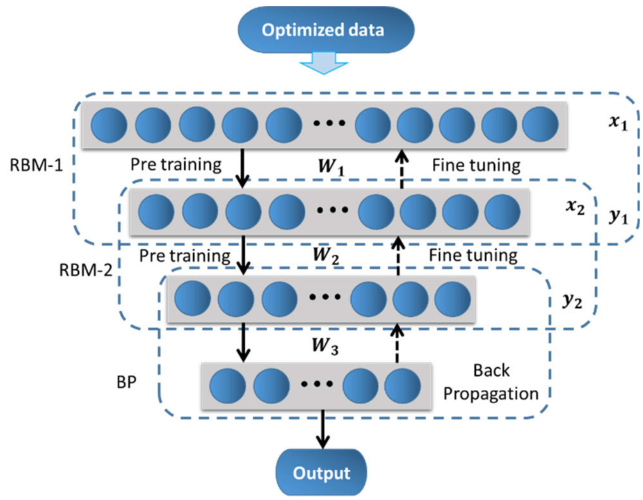


FIGURE 10. The graphical model and overview of the DBN model.

IV. EXPERIMENTAL SETTINGS AND ANALYSIS

This section includes the testing and validation performed on proposed architecture on the Egogesture and Jester dataset. On these benchmark datasets leave-one-subject-out (LOSO) cross validation test is performed to check the proposed system efficiency. From each gesture class 9 out of 10 frames were taken for training and the left out frame was taken for testing. Four experiments were performed on these two datasets for system validation. First dataset description is given and then the rest of the experiments are presented in this section with proposed system limitations, also it is compared with the other-state-of-art architectures.

A. DATASET DESCRIPTIONS

For experimentation, we used the following datasets:

1) EGOGESTURE DATASET

The egogesture [40] is a multi-model large scale dataset comprised of 2081 RGB videos with 2,953,224 frames. It includes 83 different static and dynamic gestures with 50 distinct subjects. For our system training and testing we have selected fifteen dynamic gesture classes namely: scroll hand towards left, scroll hand towards right, scroll hand downward, scroll hand upward, scroll hand backward, scroll hand forward,

TABLE 1. Confusion matrix for recognition accuracy over the Egogesture dataset.

Gestures	shtl	shtr	shd	shu	shb	shf	zift	zoft	rfc	zif	zof	sc	sd	scr	sch
shtl	9	0	0	1	0	0	0	0	0	0	0	0	0	0	0
shtr	0	10	0	0	0	0	0	0	0	0	1	0	0	0	0
shd	0	0	9	0	0	0	0	0	0	0	0	1	0	0	0
shu	0	0	0	10	0	0	0	0	1	0	0	0	0	0	0
shb	0	0	0	0	8	0	0	0	0	0	0	0	0	1	0
shf	0	0	1	0	0	9	0	0	0	0	0	0	0	0	0
zift	0	0	0	0	0	0	9	0	0	0	0	0	0	1	0
zoft	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0
rfc	0	0	0	0	1	0	0	0	9	0	0	0	0	0	0
zif	0	0	0	0	0	0	0	0	0	9	0	0	1	0	0
zof	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0
sc	0	0	0	1	0	0	0	0	0	1	0	8	0	0	0
sd	0	0	0	0	0	0	0	0	0	0	0	0	9	0	1
scr	1	0	0	0	0	0	0	0	0	0	0	0	0	9	0
sch	0	1	0	0	0	0	0	0	0	0	0	0	0	0	9
Mean Accuracy = 90.73%															

shtl = scroll hand towards left; shtr = scroll hand towards right; shd = scroll hand downward; shu = scroll hand upward; shb = scroll hand backward; shf = scroll hand forward; zift = zoom in with fists; zoft = zoom out with fists; rfc = rotate finger clockwise; zif = zoom in with finger; zof = zoom out with finger; sc = sweep circle; sd = sweep diagonal; scr = sweep cross; sch = sweep checkmark.

TABLE 2. Confusion matrix for recognition accuracy over the JESTER dataset.

Gestures	stfd	sh	ss	sl	sr	td	tu	thc	thcc	zofh	zotf	zifh	zitf	rhf	rhb
stfd	8	1	0	0	0	1	0	0	0	0	0	0	0	0	0
sh	0	9	0	0	0	0	0	0	0	0	0	0	1	0	0
ss	0	0	9	0	0	0	0	1	0	0	0	0	0	0	0
sl	0	0	0	8	0	0	0	0	1	0	0	1	0	0	0
sr	0	0	0	0	9	0	1	0	0	0	0	0	0	0	0
td	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0
tu	0	0	0	0	0	0	8	0	0	0	1	0	0	1	0
thc	0	0	0	1	0	0	0	9	0	0	0	0	0	0	0
thcc	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0
zofh	0	0	0	0	0	0	0	0	0	9	0	0	1	0	0
zotf	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0
zifh	0	0	0	0	0	0	0	0	0	1	0	9	0	0	0
zitf	0	0	0	0	0	0	0	0	0	0	0	0	9	0	1
rhf	1	0	0	0	0	0	0	0	0	0	0	0	0	9	0
rhb	0	1	0	0	0	0	0	1	0	0	0	0	0	0	8
Mean Accuracy = 89.33 %															

stfd = sliding two finger down; sh = shaking hand; ss = stop sign; sl = swiping left; sr = swipe right; td = thumb down; tu = thumb up; thc = turning hand clockwise; thcc = turning hand counterclockwise; zofh = zooming out with full hand; zotf = zooming out with two fingers; zifh = zoom in with full hand; zitf = zoom in with two fingers; rhf = rolling hand forward; rhb = rolling hand backward.

zoom in with fists, zoom out with fists, rotate finger clockwise, zoom in with finger, zoom out with finger, sweep circle, sweep diagonal, sweep cross, sweep checkmark. The gestures are collected from indoor and outdoor scenes. The dataset samples with different gesture and different backgrounds are presented in Fig. 11.

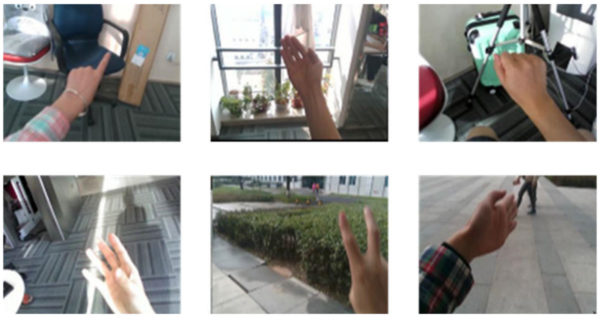


FIGURE 11. Example frames with different background of egogesture dataset.

2) JESTER DATASET

Jester dataset includes 148,092 video clips, comprises of 27 gestures of pre-defined human hand gestures collected in front of cameras [41]. The video quality of gestures is set to 100px with 12fps. 15 hand gestures are selected for system training and tested named as: Sliding two finger down, shaking hand, stop sign, swiping left, swipe right, thumb down, thumb up, turning hand clockwise, turning hand counterclockwise, zooming out with full hand, zooming out with two fingers, zoom in with full hand, zoom in with two fingers, rolling hand forward and rolling hand backward. The example gestures of jester dataset is given below in Fig 12.

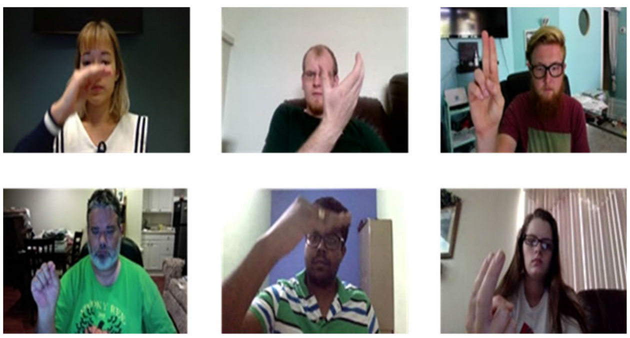


FIGURE 12. Example frames of jester dataset.

B. RECOGNITION ACCURACY

The first experiment includes the recognition accuracy of proposed system on two benchmark datasets to recognise the system efficiency and realiabilty.

1) EXPERIMENTAL SETUP

For the proposed system development, MATLAB and Google Colab were used for computations. The system comprised of 16GB of RAM, 2.6GHz with Intel (R) Core i7- 9750H CPU. To measure the proposed system accuracy, LOSO was used to analyze system efficiency. Table 1 and Table 2 demonstrates the confusion matrix with mean accuracy of Egogesture and Jester dataset respectively.

2) PERFORMANCE EVALUATION

Performance metrics are evaluated for the proposed system to demonstrate the system robustness. Precision, recall, sensitivity of the systems are calculated for both dataset gestures presented in Table 3 and 4.

TABLE 3. Evaluation metrics over Egogesture dataset.

Gestures	precision	recall	sensitivity
shtl	0.90	0.90	0.90
shtr	1.0	0.91	0.95
shd	0.90	0.90	0.90
shu	0.91	0.83	0.87
shb	0.80	0.89	0.84
shf	0.90	1.0	0.95
zift	0.90	1.0	0.95
zoft	1.0	1.0	1.0
rfc	0.90	0.90	0.90
zif	0.90	0.90	0.90
zof	1.0	0.91	0.95
sc	0.90	0.89	0.84
sd	0.90	0.90	0.90
scr	0.90	0.82	0.86
sch	0.90	0.90	0.90

TABLE 4. Evaluation metrics over JESTER dataset.

Gestures	precision	recall	sensitivity
stfd	0.80	0.89	0.84
sh	0.90	0.82	0.86
ss	0.90	1.0	0.95
sl	0.80	0.89	0.84
sr	0.90	1.0	0.95
td	1.0	0.91	0.95
tu	0.80	0.89	0.84
thc	0.90	0.82	0.86
thcc	1.0	0.91	0.95
zofh	0.90	0.90	0.90
zotf	1.0	0.91	0.95
zifh	0.90	0.90	0.90
zitf	0.90	0.82	0.86
rhf	0.90	0.90	0.90
rhb	0.80	0.89	0.84

3) COMPARISON WITH OTHER-STATE-OF-THE-ART METHODS

In this section, a comparison analysis is conducted with other state of the art algorithms on two benchmark datasets egogesture and jester. Each gesture class accuracy is compared with the other conventional algorithms namely random forest and multi-layer preceptron. Fig. 13 and 14 show the accuracies rate with conventional models on eeogesture and jester dataset respectively.

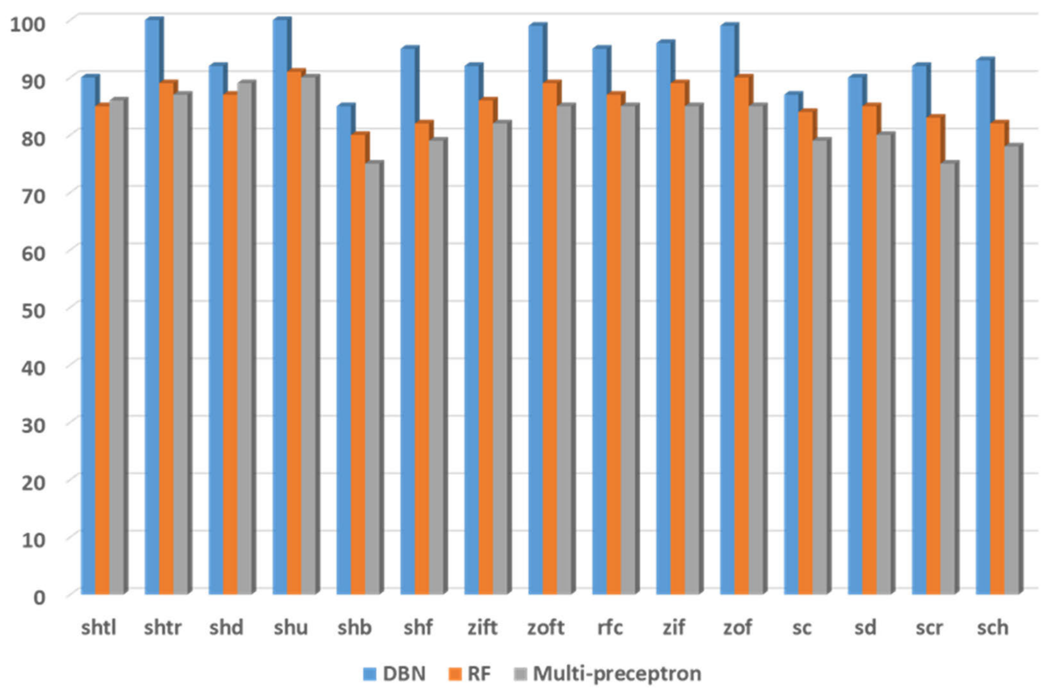


FIGURE 13. Accuracies rate over egogesture dataset with conventional model compared with proposed system.

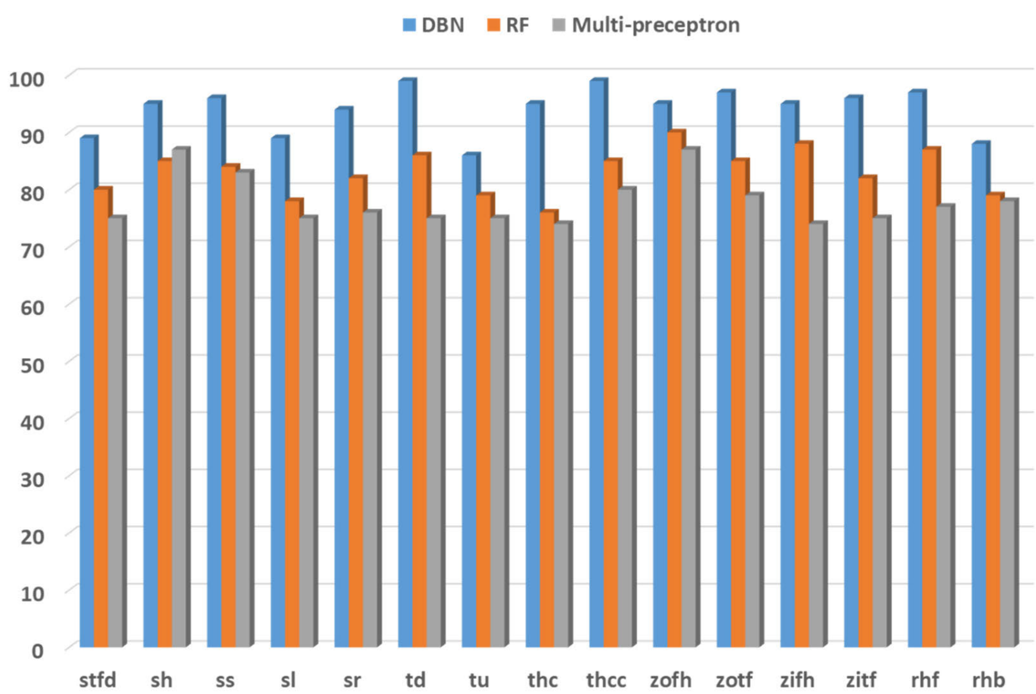


FIGURE 14. Accuracies rate over jester dataset with conventional model compared with proposed system.

4) COMPARISON WITH OTHER-STATE-OF-THE-ART TECHNIQUES

The accuracy obtained from the proposed method is compared with the other latest techniques in research area. Our system proved the best results as compared with the other state of the art method. Table 5 shows the gesture recognition accuracy over Egogesture and Jester dataset with other-state-of-the-art-methods.

TABLE 5. Comparison of gesture recognition accuracy over Egogesture and Jester dataset with other-state-of-the-art-methods.

Method	Egogesture Dataset (%)	Method	Jester Dataset (%)
Y. Liu et al. [42]	72.4	F. Fkih [44]	68.4
P. Molchanov et al. [8]	78.4	J. Yang [45]	67.9
D. Tran et al. [43]	86.4	S. Li [46]	73.55
Proposed	**90.73**	**Proposed**	**89.33**

V. CONCLUSION AND FUTURE WORK

In this paper, we have proposed the model for hand gesture recognition for medical field. Our system attained the better results considering with other conventional methods, as it detected gesture from the complex background. The proposed system detected the hand and then features were extracted, after feature optimisation feature vector was fed to classifier DBN, which proves the best accuracy results. Experimentations and evaluation done on the system provided the accuracy rate of 90.73% and 89.33% over Egogesture and jester datasets respectively.

In future, we will work comprehensively to deal with the limitations of our system. Furthermore, we will work on more hand gestures classification for medical field and also hand gesture in home automation.

Acknowlegment

The authors are thankful to Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R410), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

REFERENCES

[1] M. Nazar, M. M. Alam, E. Yafi, and M. M. Su'ud, "A systematic review of human–computer interaction and explainable artificial intelligence in healthcare with artificial intelligence techniques," *IEEE Access*, vol. 9, pp. 153316–153348, 2021.

[2] L. Nicolescu and M. T. Tudorache, "Human-computer interaction in customer service: The experience with AI chatbots—A systematic literature review," *Electronics*, vol. 11, no. 10, p. 1579, May 2022.

[3] J. Wang, R. Cheng, M. Liu, and P.-C. Liao, "Research trends of human–computer interaction studies in construction hazard recognition: A bibliometric review," *Sensors*, vol. 21, no. 18, p. 6172, Sep. 2021.

[4] Y. Sun, C. Xu, G. Li, W. Xu, J. Kong, D. Jiang, B. Tao, and D. Chen, "Intelligent human computer interaction based on non redundant EMG signal," *Alexandria Eng. J.*, vol. 59, no. 3, pp. 1149–1157, Jun. 2020.

[5] S. Nayak, B. Nagesh, A. Routray, and M. Sarma, "A human–computer interaction framework for emotion recognition through time-series thermal video sequences," *Comput. Electr. Eng.*, vol. 93, Jul. 2021, Art. no. 107280.

[6] A. Clark and D. Moodley, "A system for a hand gesture-manipulated virtual reality environment," in *Proc. Annu. Conf. South Afr. Inst. Comput. Scientists Inf. Technologists*, Sep. 2016, pp. 1–10.

[7] G. Li, H. Wu, G. Jiang, S. Xu, and H. Liu, "Dynamic gesture recognition in the Internet of Things," *IEEE Access*, vol. 7, pp. 23713–23724, 2019.

[8] P. Molchanov, X. Yang, S. Gupta, K. Kim, S. Tyree, and J. Kautz, "Online detection and classification of dynamic hand gestures with recurrent 3D convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 4207–4215.

[9] G. Plouffe and A.-M. Cretu, "Static and dynamic hand gesture recognition in depth data using dynamic time warping," *IEEE Trans. Instrum. Meas.*, vol. 65, no. 2, pp. 305–316, Feb. 2016.

[10] M. Oudah, A. Al-Naji, and J. Chahl, "Hand gesture recognition based on computer vision: A review of techniques," *J. Imag.*, vol. 6, no. 8, p. 73, Jul. 2020.

[11] N. M. Mahmoud, H. Fouad, and A. M. Soliman, "Smart healthcare solutions using the Internet of Medical Things for hand gesture recognition system," *Complex Intell. Syst.*, vol. 7, no. 3, pp. 1253–1264, Jun. 2021.

[12] N. Bargellesi, M. Carletti, A. Cenedese, G. A. Susto, and M. Terzi, "A random forest-based approach for hand gesture recognition with wireless wearable motion capture sensors," *IFAC-PapersOnLine*, vol. 52, no. 11, pp. 128–133, 2019.

[13] Y. Cho, A. Lee, J. Park, B. Ko, and N. Kim, "Enhancement of gesture recognition for contactless interface using a personalized classifier in the operating room," *Comput. Methods Programs Biomed.*, vol. 161, pp. 39–44, Jul. 2018.

[14] M. Tavakoli, C. Benussi, P. A. Lopes, L. B. Osorio, and A. T. de Almeida, "Robust hand gesture recognition with a double channel surface EMG wearable armband and SVM classifier," *Biomed. Signal Process. Control*, vol. 46, pp. 121–130, Sep. 2018.

[15] T. Zhang, H. Lin, Z. Ju, and C. Yang, "Hand gesture recognition in complex background based on convolutional pose machine and fuzzy Gaussian mixture models," *Int. J. Fuzzy Syst.*, vol. 22, no. 4, pp. 1330–1341, Jun. 2020.

[16] O. Köpüklü, A. Gunduz, N. Kose, and G. Rigoll, "Online dynamic hand gesture recognition including efficiency analysis," *IEEE Trans. Biometrics, Behav., Identity Sci.*, vol. 2, no. 2, pp. 85–97, Apr. 2020.

[17] M. Hamed, M. M. Sepehri, and T. Khatibi, "Patient's actions recognition in hospital's recovery department based on RGB-D dataset," *Multimedia Tools Appl.*, vol. 82, no. 1, pp. 24127–24154, Dec. 2022.

[18] M. Van den Bergh and L. Van Gool, "Combining RGB and ToF cameras for real-time 3D hand gesture interaction," in *Proc. IEEE Workshop Appl. Comput. Vis. (WACV)*, Jan. 2011, pp. 66–72.

[19] S. Ameur, A. B. Khalifa, and M. S. Bouhlel, "Hand-gesture-based touchless exploration of medical images with leap motion controller," in *Proc. 17th Int. Multi-Conf. Syst., Signals Devices (SSD)*, Jul. 2020, pp. 6–11.

[20] S. Ameur, A. B. Khalifa, and M. S. Bouhlel, "LeapGestureDB: A public leap motion database applied for dynamic hand gesture recognition in surgical procedures," in *Proc. 8th Int. Workshop Soft Comput. Appl.*, vol. 8, 2021, pp. 125–138.

[21] W.-J. Li, C.-Y. Hsieh, L.-F. Lin, and W.-C. Chu, "Hand gesture recognition for post-stroke rehabilitation using leap motion," in *Proc. Int. Conf. Appl. Syst. Innov. (ICASI)*, May 2017, pp. 386–388.

[22] H. P. J. Dutta, M. K. Bhuyan, D. R. Neog, K. F. MacDorman, and R. H. Laskar, "A hand gesture-operated system for rehabilitation using an end-to-end detection framework," *IEEE Trans. Artif. Intell.*, early access, Mar. 2, 2023, doi: 10.1109/TAI.2023.3251309.

[23] M. Agarwal and R. Mahajan, "Medical images contrast enhancement using quad weighted histogram equalization with Adaptive Gama Correction and homomorphic filtering," *Proc. Comput. Sci.*, vol. 115, pp. 509–517, Jan. 2017.

[24] G. Li, H. Tang, Y. Sun, J. Kong, G. Jiang, D. Jiang, B. Tao, S. Xu, and H. Liu, "Hand gesture recognition based on convolution neural network," *Cluster Comput.*, vol. 22, no. S2, pp. 2719–2729, Mar. 2019.

[25] L. Chen, J. Fu, Y. Wu, H. Li, and B. Zheng, "Hand gesture recognition using compact CNN via surface electromyography signals," *Sensors*, vol. 20, no. 3, p. 672, Jan. 2020.

[26] A. Mohanty, S. S. Rambhatla, and R. R. Sahay, "Deep gesture: Static hand gesture recognition using CNN," in *Proc. Int. Conf. Comput. Vis. Image Process.*, vol. 2, Dec. 2016, pp. 449–461.

[27] P. S. Neethu, R. Suguna, and D. Sathish, "An efficient method for human hand gesture detection and recognition using deep learning convolutional neural networks," *Soft Comput.*, vol. 24, no. 20, pp. 15239–15248, Oct. 2020.

[28] Y. Kim and B. Toomajian, "Hand gesture recognition using micro-Doppler signatures with convolutional neural network," *IEEE Access*, vol. 4, pp. 7125–7130, 2016.

[29] M. Al-Hammadi, G. Muhammad, W. Abdul, M. Alsulaiman, and M. S. Hossain, "Hand gesture recognition using 3D-CNN model," *IEEE Consum. Electron. Mag.*, vol. 9, no. 1, pp. 95–101, Jan. 2020.

[30] A. R. Asif, A. Waris, S. O. Gilani, M. Jamil, H. Ashraf, M. Shafique, and I. K. Niazi, "Performance evaluation of convolutional neural network for hand gesture recognition using EMG," *Sensors*, vol. 20, no. 6, p. 1642, Mar. 2020.

[31] P. M. Yanik, J. Manganelli, J. Merino, A. L. Threatt, J. O. Brooks, K. E. Green, and I. D. Walker, "Use of Kinect depth data and growing neural gas for gesture based robot control," in *Proc. 6th Int. Conf. Pervasive Comput. Technol. Healthcare (PervasiveHealth) Workshops*, May 2012, pp. 283–290.
[32] J. Botzheim and N. Kubota, "Growing neural gas for information extraction in gesture recognition and reproduction of robot partners," in *Proc. Int. Symp. Micro-NanoMechatronics Human Sci. (MHS)*, Nov. 2012, pp. 149–154.
[33] H. Ansar, A. Jalal, M. Gochoo, and K. Kim, "Hand gesture recognition based on auto-landmark localization and reweighted genetic algorithm for healthcare muscle activities," *Sustainability*, vol. 13, no. 5, p. 2961, Mar. 2021.
[34] A. Dayal, N. Paluru, L. R. Cenkeramaddi, J. Soumya, and P. K. Yalavarthy, "Design and implementation of deep learning based contactless authentication system using hand gestures," *Electronics*, vol. 10, no. 2, p. 182, Jan. 2021.
[35] A. Mujahid, M. J. Awan, A. Yasin, M. A. Mohammed, R. Damaševičius, R. Maskeliūnas, and K. H. Abdulkareem, "Real-time hand gesture recognition based on deep learning YOLOv3 model," *Appl. Sci.*, vol. 11, no. 9, p. 4164, May 2021.
[36] T. Ozcan and A. Basturk, "Transfer learning-based convolutional neural networks with heuristic optimization for hand gesture recognition," *Neural Comput. Appl.*, vol. 31, no. 12, pp. 8955–8970, Dec. 2019.
[37] Y. Wang, Z. Pan, X. Yuan, C. Yang, and W. Gui, "A novel deep learning based fault diagnosis approach for chemical process with extended deep belief network," *ISA Trans.*, vol. 96, pp. 457–467, Jan. 2020.
[38] Y. Zhang, P. Li, and X. Wang, "Intrusion detection for IoT based on improved genetic algorithm and deep belief network," *IEEE Access*, vol. 7, pp. 31711–31722, 2019.
[39] M. M. Hassan, M. G. R. Alam, M. Z. Uddin, S. Huda, A. Almogren, and G. Fortino, "Human emotion recognition using deep belief network architecture," *Inf. Fusion*, vol. 51, pp. 10–18, Nov. 2019.
[40] T. Chalasani and A. Smolic, "Simultaneous segmentation and recognition: Towards more accurate ego gesture recognition," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 4367–4375.
[41] J. Materzynska, G. Berger, I. Bax, and R. Memisevic, "The jester dataset: A large-scale video dataset of human gestures," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 2874–2882.
[42] Y. Liu, D. Jiang, H. Duan, Y. Sun, G. Li, B. Tao, J. Yun, Y. Liu, and B. Chen, "Dynamic gesture recognition algorithm based on 3D convolutional neural network," *Comput. Intell. Neurosci.*, vol. 2021, pp. 1–12, Aug. 2021.
[43] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Learning spatiotemporal features with 3D convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 4489–4497.
[44] F. Fkih, "Similarity measures for collaborative filtering-based recommender systems: Review and experimental comparison," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 34, no. 9, pp. 7645–7669, Oct. 2022.
[45] J. Yang, X. Dong, L. Liu, C. Zhang, J. Shen, and D. Yu, "Recurring the transformer for video action recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 14043–14053.
[46] S. Li, A. Aich, S. Zhu, M. S. Asif, C. Song, A. K. Roy-Chowdhury, and S. V. Krishnamurthy, "Adversarial attacks on black box video classifiers: Leveraging the power of geometric transformations," in *Proc. 35th Conf. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 2085–2096.

MOHAMMED ALONAZI received the B.Sc. degree in computer science from King Saud University, Saudi Arabia, in 2008, the M.Sc. degree in computer science from the Florida Institute of Technology, Melbourne, FL, USA, in 2015, and the Ph.D. degree in informatics from the University of Sussex, U.K., in 2019. He is currently an Assistant Professor with the Department of Information Systems, College of Computer Engineering and Sciences, Prince Sattam bin Abdulaziz University, Al-Kharj, Saudi Arabia. His research interests include human–computer interaction, UX/UI, digital transformation, cyber security, and machine learning.

HIRA ANSAR received the B.S. degree (Hons.) in computer science from Fatima Jinnah Women, Rawalpindi, Pakistan. She is currently pursuing the M.S. degree in computer science with Air University, Islamabad, Pakistan. Her research interests include artificial intelligence, computer vision, machine learning algorithms, deep learning, image and video processing, and intelligent systems. She received the Silver Medal for the B.S. degree.

NAIF AL MUDAWI received the master's degree in computer science from Australian La Trobe University, in 2011, and the Ph.D. degree from the College of Engineering and Informatics, University of Sussex, Brighton, U.K., in 2018. He is currently an Assistant Professor with the Department of Computer Science and Information System, Najran University. He has many published research and scientific papers in many prestigious journals in various disciplines of computer science. He was a member of the Australian Computer Science Committee.

SAUD S. ALOTAIBI received the Bachelor of Computer Science degree from King Abdul Aziz University, in 2000, the master's degree in computer science from King Fahd University, Dhahran, in May 2008, and the Ph.D. degree in computer science from Colorado State University, Fort Collins, CO, USA, under the supervision of Dr. Charles Anderson, in August 2015. He started his career as an Assistant Lecturer with Umm Al-Qura University, Makkah, Saudi Arabia, in July 2001. He was a Deputy of the IT-Center for E-Government and Application Services, Umm Al-Qura University, in January 2009. From 2015 to 2018, he was with the Deanship of Information Technology to improve the IT services that are provided to Umm Al-Qura University. He is currently an Assistant Professor of computer science with Umm Al-Qura University. He is currently the Vice Dean for academic affairs with the Computer and Information College. His current research interests include AI, machine learning, natural language processing, neural computing, the IoT, knowledge representation, smart cities, wireless, and sensors.

NOUF ABDULLAH ALMUJALLY received the Ph.D. degree in computer science from the University of Warwick, U.K. She is currently an Assistant Professor of computer science with the Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University (PNU), Riyadh, Saudi Arabia. Her research interests include human–computer interaction (HCI), artificial intelligence (AI), machine learning, deep learning, and computer-based applications.

ABDULWAHAB ALAZEB received the B.S. degree in computer science from King Khalid University, Abha, Saudi Arabia, in 2007, the M.S. degree in computer science from the Department of Computer Science, University of Colorado Denver, USA, in 2014, and the Ph.D. degree in cybersecurity from the University of Arkansas, USA, in 2021. He is currently an Assistant Professor with the Department of Computer Science and Information System, Najran University. His research interests include cybersecurity, cloud and edge computing security, machine learning, and the Internet of Things. He received the Graduate Certificate in cybersecurity from the University of Arkansas.

AHMAD JALAL received the Ph.D. degree from the Department of Biomedical Engineering, Kyung Hee University, Republic of Korea. He was a Postdoctoral Research Fellow with POSTECH. He is currently an Associate Professor with the Department of Computer Science and Engineering, Air University, Pakistan. His research interests include multimedia contents, artificial intelligence, and machine learning.

JAEKWANG KIM (Member, IEEE) received the B.S., M.S., and Ph.D. degrees from Sungkyunkwan University, Republic of Korea, in 2004, 2006, and 2014, respectively. He has been a Staff of University Bible Fellowship, Republic of Korea, since 2014. He is currently an Assistant Professor with Sungkyunkwan University. He is also a Born Again Christian. He is affiliated with multiple departments with Sungkyunkwan University, including the School of Convergence, the Department of Computing, the Department of Applied Data Science, the Department of Applied Artificial Intelligence, and the Convergence Program for Social Innovation as well as the AI Research Institute. His research interests include artificial intelligence, with a focus on recommendation algorithms, intelligent network and security, and AI core and applied algorithms in various domains.

MOOHONG MIN received the B.S. and M.S. degrees from Sungkyunkwan University, in 2010 and 2012, respectively, and the Ph.D. degree from Korea University. He is currently an Assistant Professor with Sungkyunkwan University, Republic of Korea. He is affiliated with multiple departments with Sungkyunkwan University, including the Department of Computer Education and the Social Innovation Convergence Program. His research interests include anomaly detection techniques using machine learning and advanced study of information security.