Automated Gesture Recognition Using African Vulture Optimization with Deep Learning for Visually Impaired People on Sensory Modality Data

Mashael Maashi[1,2], Mohammed Abdullah Al-Hagery[3], Mohammed Rizwanullah[4,*] and Azza Elneil Osman[4]

[1]Department of Software Engineering, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia
[2]King Salman Center for Disability Research, Riyadh, Saudi Arabia
[3]Department of Computer Science, College of Computer, Qassim University, Saudi Arabia
[4]Department of Computer and Self Development, Preparatory Year Deanship, Prince Sattam Bin Abdulaziz University, Al Kharj, Saudi Arabia

Correspondence to:
Mohammed Rizwanullah*, e-mail: r.mohammed@psau.edu.sa

ABSTRACT

Gesture recognition for visually impaired persons (VIPs) is a useful technology for enhancing their communications and increasing accessibility. It is vital to understand the specific needs and challenges faced by VIPs when planning a gesture recognition model. But, typical gesture recognition methods frequently depend on the visual input (for instance, cameras); it can be vital to discover other sensory modalities for input. The deep learning (DL)-based gesture recognition method is effective for the interaction of VIPs with their devices. It offers a further intuitive and natural way of relating with technology, creating it more available for everybody. Therefore, this study presents an African Vulture Optimization with Deep Learning-based Gesture Recognition for Visually Impaired People on Sensory Modality Data (AVODL-GRSMD) technique. The AVODL-GRSMD technique mainly focuses on the utilization of the DL model with hyperparameter tuning strategy for a productive and accurate gesture detection and classification process. The AVODL-GRSMD technique utilizes the primary data preprocessing stage to normalize the input sensor data. The AVODL-GRSMD technique uses a multi-head attention-based bidirectional gated recurrent unit (MHA-BGRU) method for accurate gesture recognition. Finally, the hyperparameter optimization of the MHA-BGRU method can be performed by the use of African Vulture Optimization with Deep Learning (AVO) approach. A series of simulation analyses were performed to demonstrate the superior performance of the AVODL-GRSMD technique. The experimental values demonstrate the better recognition rate of the AVODL-GRSMD technique compared to that of the state-of-the-art models.

KEYWORDS

African vulture optimization algorithm, deep learning, gesture recognition, visually impaired person, hyperparameter tuning

INTRODUCTION

Human–computer interaction (HCI) refers to a regulated loop process where data sent by devices and humans should be considered whole (Padmavathi, 2021). This interaction loop is an integration of feedforwards and feedback. Meeting the challenges of blind persons in accessing contextual and graphical data is challenging (Agarwal and Das, 2023). Gesture detection has a considerable amount of realistic applications that include remote teaching guidance, virtual reality/augmented reality (VR/AR) games, and interaction with remote automation equipment operations, remote health care, and intelligent vehicles (Gorobets et al., 2022). Moreover, gesture detection grabs special interest to understand the ways to enhance the standard of life of people with hearing disorders, and gesture detection is utilized in translation tasks and sign language recognition for deaf persons

(Gangrade and Bharti, 2023). Several research studies on gesture detection were carried out, and remarkable progression has been made in this field.

Gesture detection is classified into dynamic gesture detection and static gesture detection (Li et al., 2022). Static gesture detection needs organizing and learning the gestures' spatial features without concerning the temporal features (Ryumin et al., 2023). Conversely, dynamic gesture detection needs to consider both the temporal and spatial features of the gesture since it varies over period. Hence, dynamic gesture detection is highly sophisticated compared to static gesture detection; however, the utilization of dynamic gestures is broader (Sahana et al., 2022). This study offers a lightweight gesture action detection network for real-time HCI and control. However, the potentiality of potential

gesture detection remains an unsolved issue because of the variances in the syntactic and semantic framework of gestures (Pandey, 2023). At present, fully automatic methods for detecting various dynamic gestures do not exist (Gangrade and Bharti, 2023). Devising these methods needs deep semantic analyses that can be conducted at a superficial level owing to the limitation of text analysis approaches and knowledge bases and can attain word level detection. Thus the presented approach has few reference important for the sign language detection of persons with hearing disorders (Varsha and Nair, 2021).

This study presents an African Vulture Optimization with Deep Learning-based Gesture Recognition for Visually Impaired People on Sensory Modality Data (AVODL-GRSMD) technique. The AVODL-GRSMD technique utilizes the primary data preprocessing stage to normalize the input sensor data. The AVODL-GRSMD technique uses a multi-head attention-based bidirectional gated recurrent unit (MHA-BGRU) method for accurate gesture recognition. Finally, the hyperparameter optimization of the MHA-BGRU technique is performed by the use of the AVO method. A series of simulation analyses were conducted to validate the improved performance of the AVODL-GRSMD technique.

RELATED STUDIES

The authors in Adeel et al. (2022) define a gesture-based confidence assessment (GCA) method for hand gesture recognition (HGR) for identifying the state of mind based on the hand actions in an interview as a context. This method is also valuable for visually impaired persons (VIPs) when conducting or performing an interview. Previously, there was without work completed to identify a person's state of mind utilizing HGR. This method was dependent upon a conventional neural network (CNN) with long short-term memory networks (LSTM) for capturing the temporal data. In Zhang and Zeng (2022), touch gestures were predictable by a trained radial basis function (RBF) network, but integrated gestures can be demonstrated by Petrinet, which establishes a logic, timing, and spatial relationship model. As a result, the Braille input regarding multi-touch gesture recognition was executed.

Deepa et al. (2023) presented a structure that utilizes CNN for HGR. The gesture sign has been verified and exposed to binarization, but the image was divided into background and foreground. Contours can be identified in the binarized image. Feature extraction has been completed utilizing the SIFT technique. The feature extraction can be used by the CNN technique for recognizing the HGR. The HGR was provided as an outcome from the text format. Can et al. (2021) proposed a DL-CNN approach which categorizes hand gestures efficiently in the investigation of near-infrared and color natural images. This paper presents a novel deep learning (DL) technique dependent upon CNN for recognizing hand gestures, enhancing the rate of recognition, testing, and training time.

Al-Hammadi et al. (2020) presented an effective diffusion-convolutional neural networks (DCNN) algorithm for HGR. The presented technique utilized transfer learning (TL) to beat the lack of huge labeled hand gesture database.

Kraljević et al. (2020) examined a smart home automatization method specially planned for providing real-time sign language recognition. A new hierarchical system can be projected comprising resource-and-time-aware elements—a wake-up element and better performance sign recognition element dependent upon the Conv3D network. Lahiani and Neji (2018) proposed a static HGR method for mobile devices by integrating the histogram of orientated gradients (HOG) and local binary pattern (LBP) features that correctly identify hand poses.

THE PROPOSED MODEL

This study concentrates on the development of an automated gesture recognition tool named the AVODL-GRSMD technique for visually impaired people based on sensory modality data. The AVODL-GRSMD technique exploits the DL model with hyperparameter tuning strategy for effectual and accurate gesture detection and classification process. The AVODL-GRSMD technique follows three major processes, namely data preprocessing, MHA-BGRU recognition, and AVO-based hyperparameter tuning. Figure 1 demonstrates the workflow of the AVODL-GRSMD system.

Data preprocessing

To preprocess the input data, data normalization is adopted. The data recorded by wearable sensors were normalized and cleaned for achieving suitable and consistent data for training a recognition element. Primarily, the imputation process can set the missing values of sensor databases with the linear interpolation method. Next, the noises can be removed with median filtering and a third-order low-pass Butterworth filter with a 20 Hz cutoff frequency. A normalized model converts all the sensor data with mean and standard derivation.

Gesture recognition using the MHA-BGRU technique

At this stage, the preprocessed input can be passed into the MHA-BGRU method for gesture recognition. Recurrent neural network (RNN) is specialized to process sequence data, namely audio, time series, and text, different from CNN, focusing on the spatial features of the input (Bao et al., 2022). In general, RNN conducts a similar computation process cyclically on all the segments of the sequence, and the subsequent output depends on prior calculation. From the network architecture, it involves memory for storing the hidden internal state h_t that is evaluated by the input x_t and prior hidden layer (HL) h_{t-1}, such that

$$h_t = f_W\left(W_{xh}x_t + W_{hh}h_{t-1} + b_h\right). \tag{1}$$

In Eq. (1), W_{hh} denotes the weight of the hidden-to-HL, and W_{yh} indicates the weight of hidden-to-output; f_w denotes the HL function, namely tanh activation function with W
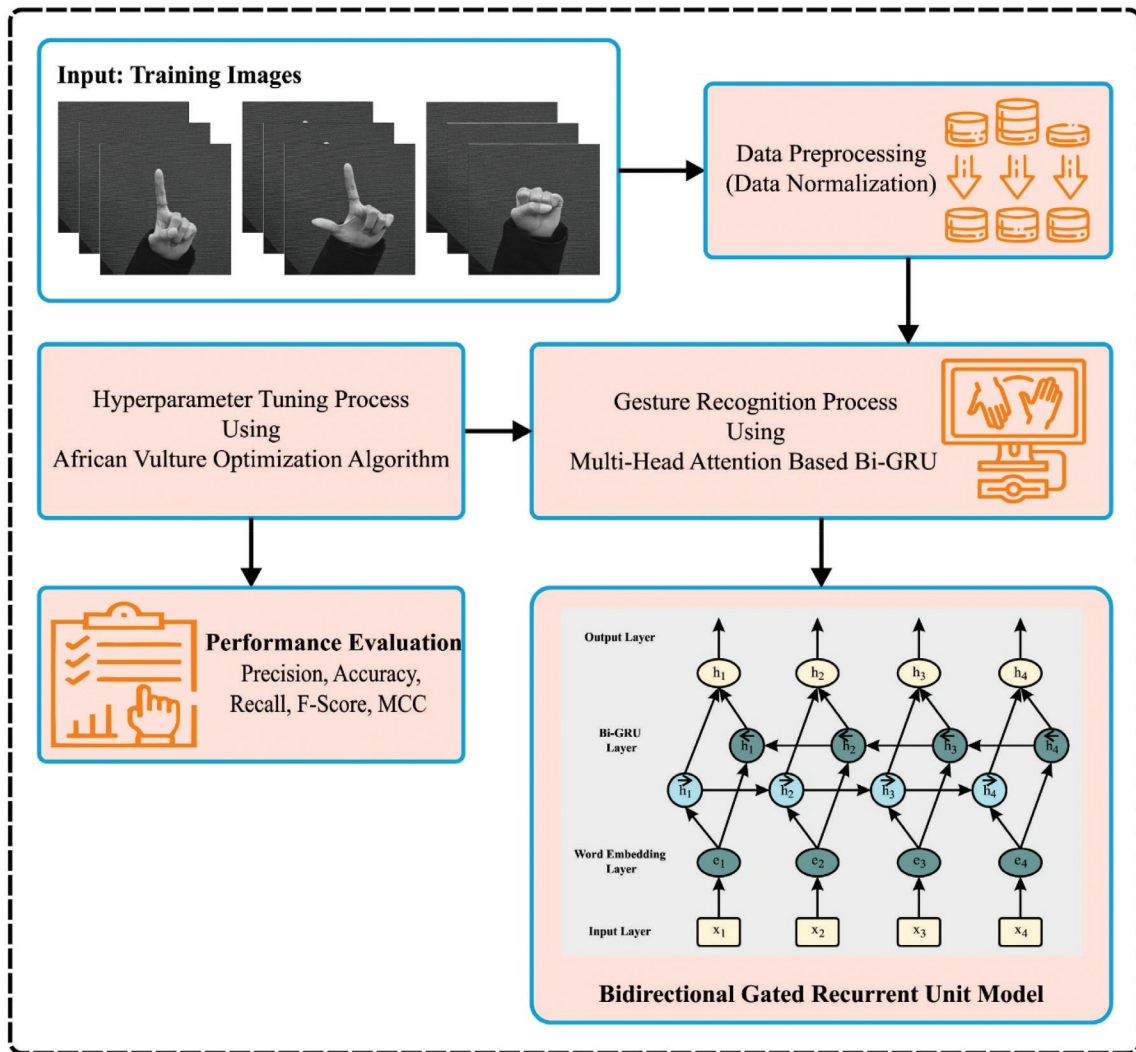
Figure 1: Workflow of the AVODL-GRSMD system. Abbreviation: AVODL-GRSM, African Vulture Optimization with Deep Learning based Gesture Recognition for Visually Impaired People on Sensory Modality Data.

parameter shared across time (viz., W_{xh} specifies the weight of the input-HL); b represents the corresponding bias vector, and the predicted output is as follows:

$$y_t = W_{yh}h_t + b_y. \tag{2}$$

But this typical architecture faces problems of vanishing gradients and exploding weights on long-term sequences.

The bidirectional RNN structure allows the output layer to receive future and past data for all the points in the input series. More accurately, a forward RNN learns from prior information, whereas reverse RNN learns from future information in such a way that all the time steps make optimum usage of lower- and upper-relevant data. In addition, both outputs are spliced together as the concluding output of bidirectional recurrent neural network (BiRNN).

Given that, BiGRU is a BiRNN that exploits the GRU for all the hidden nodes. BiGRU splits GRU neurons into backward and forward layers that match with negative and positive time directions, correspondingly.

The existing statement of the HL of BiGRU can be defined by the existing input x_t, the HL statement output of backward layer $\overleftarrow{h_{t-1}}$, and the forward layer $\overrightarrow{h_{t-1}}$. Meanwhile, BiGRU

is considered as two single GRUs, the HL state of BiGRU at t time is attained by the weighted amount of $\overrightarrow{h_{t-1}}$ and $\overleftarrow{h_{t-1}}$ that is expressed below:

$$\overrightarrow{h_t} = GRU\left(x_t, \overrightarrow{h_{t-1}}\right) \tag{3}$$

$$\overleftarrow{h_t} = GRU\left(x_t, \overleftarrow{h_{t-1}},\right)b \tag{4}$$

$$h = w\overrightarrow{h_t} + v_t\overleftarrow{h_t} + b_t. \tag{5}$$

Briefly, BiGRU allows modeling the possible relationship between future and historical ship trajectory status with the existing state, thus raising prediction accuracy.

The attention-based mechanism was devised in the field of image detection, and now it is utilized instead of RNN in the field of machine translation. The attention-based module highlighted the curial influencing factors. By allocating a weight to all the elements in the input sequence, thus increasing accuracy of the model:

$$f(x_i, y) = (W_1*x_i, W_2*y) \tag{6}$$

$$Attention = \sum_{i=1}^{n} softmax(f(x_i, y))*x_i, \tag{7}$$

where x_i signifies the input series. It can be mapped in the [0,1] interval by the normalized exponential function that can be "weight." Moreover, Dot product attention refers to the weighted combination of x_i.

The multi-head attention (MHA) module emerges as the situation requires, with the attention-related module being extensively utilized in image and natural language processing (NLP) tasks. Furthermore, all the iteration's linear conversion parameters W for, K, and V seems to be unique; they are not shared. MHA can be utilized for processing the information from the BiGRU output layer instead of applying average or maximum pooling, as follows:

$$Attention\ (Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \qquad (8)$$

$$\begin{aligned} Multi-Head\ (Q,K,V) \\ = (head_1 \oplus head_2 \oplus head_3 \oplus \cdots \oplus head_h)W^0. \end{aligned} \qquad (9)$$

Therefore, the MHA module, a fusion of multi-attention-based mechanisms, is considered a weighting system for information that could allocate weight to the HL of BiGRU so that they effectively apply data sources while making predictions.

Hyperparameter tuning using the AVO algorithm

Finally, the AVO algorithm can be applied for the optimal hyperparameter adjustment of the MHA-BGRU model. The process involved in the AVO algorithm is elaborated as follows (Liu et al., 2023). The fitness value for every answer was evaluated after forming the initial population, the better answer of the initial Vulture group and the better answer of the second Vulture group is defined, and the following equation can choose other answers.

$$R(i) = \begin{cases} BestVulture_1\ if\ p_i = L_1 \\ BestVulture_2\ if\ p_j = L_2 \end{cases}. \qquad (10)$$

In Eq. (10), the parameters L_1 and L_2 should be initialized with values within [0,1] before the search process, and the sum of both variables was equivalent to one as follows:

$$p_i = \frac{F_i}{\sum_{i=1}^{n} F_i} \qquad (11)$$

Equation (13) is used for mathematical modeling of these behaviors. It is used for transition from ERP to ETp that is stimulated by the satiety speed or hunger of Vultures.

$$T = h \times \left(sin^w\left(\frac{\pi}{2} \times \frac{iter_i}{max\_iter}\right) + cos\left(\frac{\pi}{2} \times \frac{iter_i}{max\_iter}\right) - 1\right)(12)$$

$$F = (2 \times rand_1 + 1) \times z \times \left(1 - \frac{iter_i}{max\_iter}\right) + t \qquad (13)$$

In Eqs. (12) and (13), F denotes that the agent is satisfied, $iter_i$ shows the existing amount of iteration, max_iter

represents the overall amount of iterations, and z represents the randomly generated value within $[-1,1]$ which changes all the iterations, h denotes the random integer within $[-2, 2]$. $rand_1$ represents the random integer within [0,1], and w denotes the parameter with constant number. If $|F| > 1$ the agent looks for food in dissimilar regions, and AVO algorithm entered the ERP. If $|F| < 1$ AVO algorithm enters the ETp, and agent forage in the neighborhood of solution.

Here, the variable P_1, which ranges from 0 to 1, is used for choosing two dissimilar approaches. Before the search operation, these parameters should be valued. A random integer within [0,1] is generated for selecting every strategy in the ERP rand. If $\geq P_1$, Eq. (15) is exploited. If $rand_{p_1} < P_1$, Eq. (16) is applied:

$$P1(i+1)\begin{cases} Eq.\ (15)\ if\ p_1 \geq rand_{p_1} \\ Eq.\ (16)\ if\ p_1 \geq rand_{p_1} \end{cases} \qquad (14)$$

$$P1(i+1) = R(i) - D(i) \times F \qquad (15)$$

$$D(i) = |X \times R(i) - P1(i)|. \qquad (16)$$

In Eq. (14), $P1(i+1)$ denotes the agent's position vector in the following iteration, and F indicates the agent's satiation rate. In Eq. (16), $R(i)$ denotes the better agent. Moreover, X denotes the agent randomly moving to secure food from other agents and can be provided by $X = 2 \times rand$, where $rand$ denotes the random integer within [0, 1]. $P1(i)$ denotes the existing position vector of the Vultures.

$$P1(i+1) = R(i) - F + rand_2 \times (UB - LB) \times rand_3 + LB), \quad (17)$$

where $rand_2$ has a random integer within [0,1], and $rand_3$ has taken a number closer to 1.

The fourth stage: exploitation

If $|F| < 1$, AVO algorithm entered the ETp that also has stages where dual dissimilar strategies are applied. The selection degree of all the strategies in every internal stage is defined by two parameters, such as P_2 and P_3. In the first stage, the P_2 variable is used for selecting the strategy; in the second stage, the P_3 parameter is used for selecting the strategy. Both parameters should be set to 0 and 1 before carrying out the searching process.

The AVO algorithm enters the initial phase in the ETp if $|F|$ is between 1 and 0.5. Initially, two dissimilar approaches of turn flight and siege combat are implemented. P_2 represents the selection of all the strategies that should be estimated beforehand by implementing the search process, and the value should be between 0 and 1. Initially, $rand$, a random integer within [0,1], is produced. If the number is $\geq P_2$, the Siegefight approach is gradually implemented. But if this random number is $< P_2$, then rotary flight strategy is implemented.

$$P1(i+1)\begin{cases} Eq.\ (20)\ if\ p_2 \geq rand_{p_2} \\ Eq.\ (23)\ if\ p_2 \geq rand_{p_2} \end{cases}. \qquad (18)$$

If $|F| \geq 0.5$, then the agent was comparatively full and had sufficient energy. If the agent congregates on a food source,

then it causes intense conflict over food. The weak agent tries to take and exhaust food from healthy agents by collecting around the healthy agent.

$$P1(i+1) = D(i) \times (F + rand_4) - d(t) \quad (19)$$

$$d(t) = R(i) - P1(i), \quad (20)$$

where $D(i)$ is evaluated using Eq. (16), and F denotes the satiety of the agent; $rand_4$ denotes the random integer within [0,1], which is used for increasing the randomness factor. $R(i)$ is the better agent of the second category chosen by means of Eq. (10). In the existing iteration, $P1(i)$ indicates the agent's existing position vector, where the distance between agent and better agent in the second category is attained.

Agent frequently implements a circling flight. A spiral is generated between each agent and top two agents, and it can be discussed as follows:

$$S_1 = R(i) \times \left(\frac{rand_5 \times P1(i)}{2\pi} \right) \times \cos(P1(i)),$$

$$S_2 = R(i) \times \left(\frac{rand_6 \times P1(i)}{2\pi} \right) \times \sin(P1(i)) \quad (21)$$

$$P1(i+1) = R(i) - (S_1 + S_2). \quad (22)$$

$R(i)$ indicates the position vector of the two better agents in the existing iteration. *Cos* and *sin* represent the sin and cos functions correspondingly. $rand_5$ and $rand_6$ show randomly generated numbers between 0 and 1. S_1 and S_2 are attained using Eq. (22).

Exploitation (second stage)

Here, *ETp*, the movement of two agents, collects different kinds of agents over the food sources, and aggressive and encirclement fighting is performed for finding food. If $|F|$ < 0.5, this step of the method was implemented. Initially, $rand_{p_3}$ is produced that is randomly generated within [0,1]. If $rand_{p_3} \geq P_3$, the strategy was to collect different kinds of agents on the food source. Or else, if the value is <P_3, then siege fight offensive strategy was implemented.

$$P1(i+1) = \begin{cases} Eq.(26) \ if \ p_3 \geq rand_{p_3} \\ Eq. (27) \ if \ p_3 \geq rand_{p_3} \end{cases} \quad (23)$$

$$A_1 = BestVulture \ (i) - \frac{BestVulture_1(i) \times p1(i)}{BestVulture_1(i) - p1(i)^2} \times F,$$

$$A_2 = BestVulture(i) - \frac{BestVulture_2(i) \times p1(i)}{BestVulture_2(i) - p1(i)^2} \times F, \quad (24)$$

where $BestVulture_1(i)$ and $BestVulture_2(i)$ denote the better agent from initial and second category in the existing iteration, whereas F denotes the agent's satiety, and $P1(i)$ shows the existing vector position of the agent.

$$P1(i+1) = \frac{A_1 + A_2}{2}. \quad (25)$$

Lastly, the aggregation of each agent is performed using Eq. (26), where A_1 and A_2 are attained using Eq. (24), and $P1(i+1)$ indicates agent position vector in the following iteration. Figure 2 represents the flowchart of the AVO algorithm.

If $|F|$ < 5, head agent becomes hungry and weaker and doesn't have sufficient energy to fight other agents. At the same time, other agents turn out to be aggressive while finding food. They move in dissimilar directions toward the head of agent.

$$P1(i+1) = R(i) - |d(t)| \times F \times Levy(d), \quad (26)$$

where $d(t)$ characterizes the distance of the agent to the better agent of the second category that is evaluated using Eq. (20). Levy flight (LF) designs are employed for higher performance of african-vulture optimization algorithm (AVOA) in Eq. (26), and LF was recognized and utilized in several MH approaches. *LF* is computed utilizing Eq. (27).

$$LF(x) = 0.01 \times \frac{u \times \sigma}{|v|^{\frac{1}{p}}}, \sigma = \left(\frac{\Gamma(1+\beta) \times \sin\left(\frac{\pi\beta}{2}\right)}{\Gamma(1+\beta2) \times \beta \times 2^{\left(\frac{\beta-1}{2}\right)}} \right)^{\frac{1}{p}} \cdot (27)$$

Fitness choice is a critical aspect of the AV system. An encoded result can be utilized to evaluate the aptitude of candidate results. Recently, the accuracy value is key condition employed for scheming a fitness function.
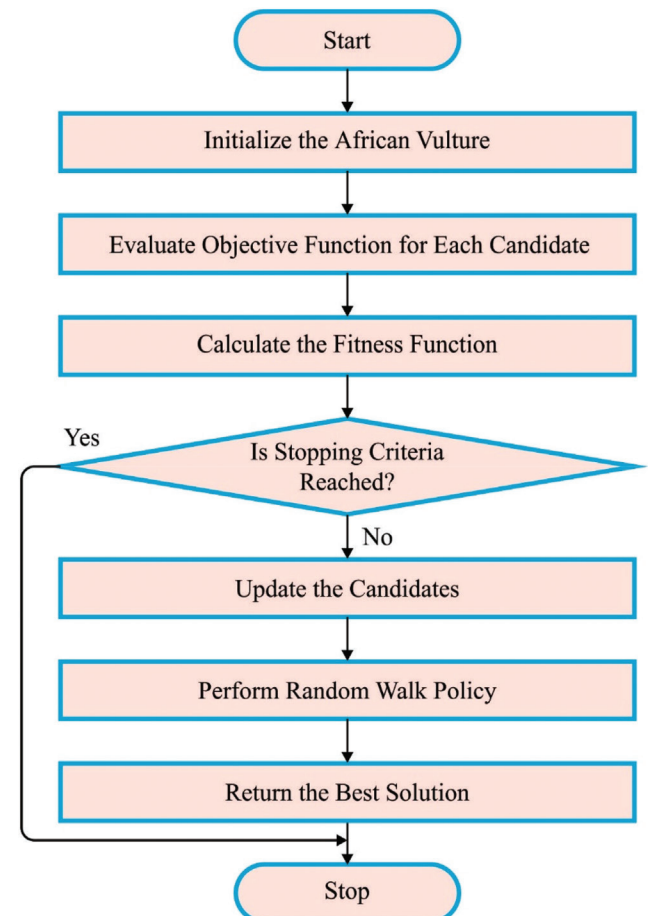

Figure 2: Flowchart of the AVO algorithm.

$$Fitness = \max(P) \tag{28}$$

$$P = \frac{TP}{TP + FP}, \tag{29}$$

in which *TP* demonstrates the true positive and *FP* signifies the false positive values.

Table 1: Details on two databases.

Class	Labels	Dataset	
		UCI HAR	USC HAD
		No. of samples	
Walking	C-1	1722	8476
Walking upstairs	C-2	1544	4709
Walking downstairs	C-3	1406	4382
Sitting	C-4	1777	5810
Standing	C-5	1906	5240
Laying/sleeping	C-6	1944	8331
Total number of samples		10299	36948

RESULTS AND DISCUSSION

In this section, the gesture recognition results of the AVODL-GRSM technique are validated on two datasets: UCI HAR dataset (UCI HAR) and USC HAD dataset (USC HAD), as illustrated in Table 1.

Figure 3 demonstrates the classifier results of the AVODL-GRSM technique on the UCI HAR dataset. Figure 3a and b portray the confusion matrix rendered by the AVODL-GRSM method on 70:30 of TRP/TSP. The figure specifies that the AVODL-GRSM technique has identified and classified all six class labels accurately. Likewise, Figure 3c shows the PR analysis of the AVODL-GRSM approach. The figures pointed out that the AVODL-GRSM approach has acquired maximum PR performance under six classes. Eventually, Figure 3d illustrates the receiver operating characteristic (ROC) investigation of the AVODL-GRSM model. The results portrayed that the AVODL-GRSM approach has proficient outcomes with maximum ROC values under six class labels.
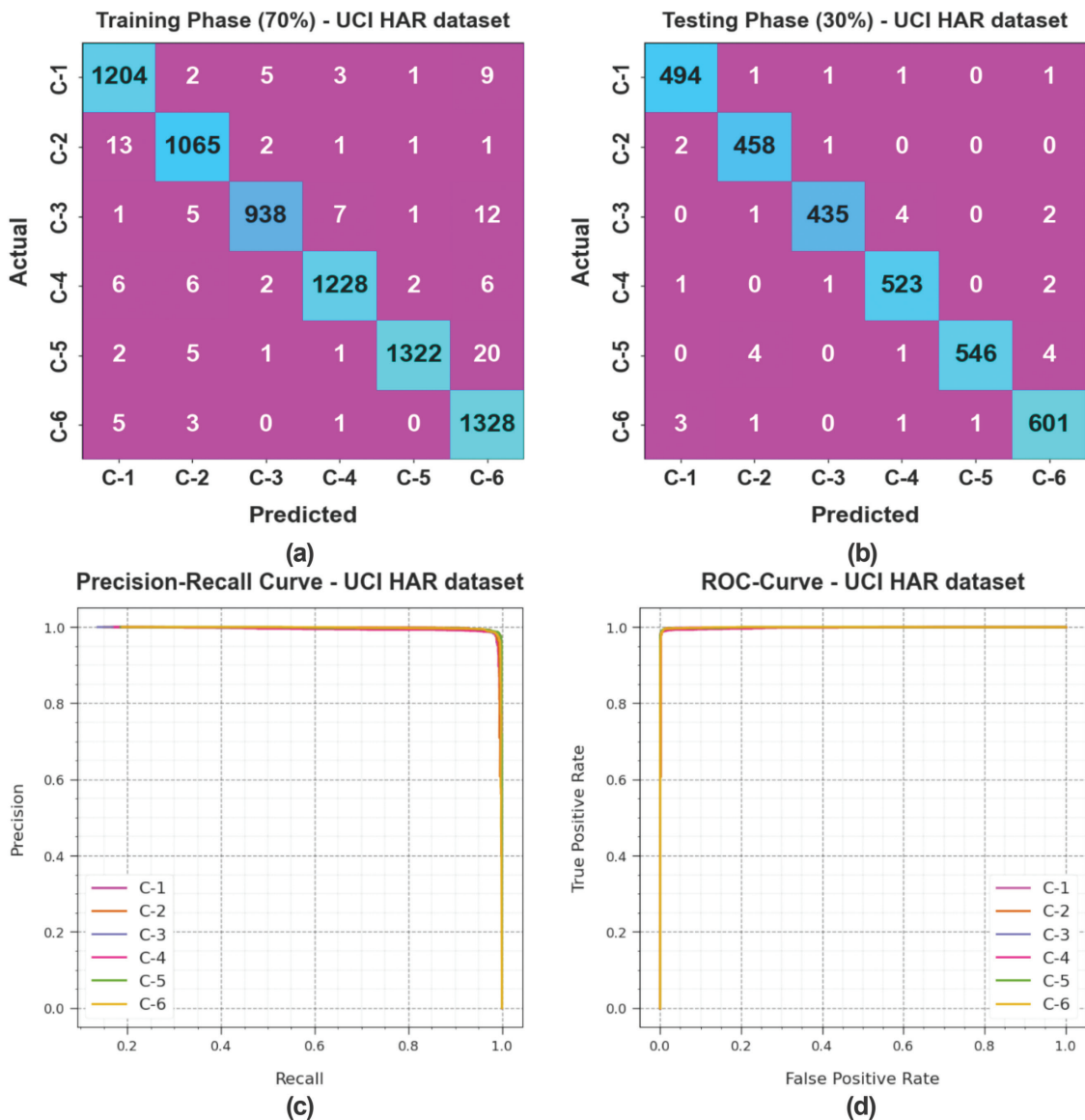

Figure 3: Classifier result on the UCI HAR dataset. (a and b) Confusion matrices, (c) PR curve, and (d) ROC curve.

In Table 2, a brief recognition result of the AVODL-GRSM technique is clearly portrayed on the UCI HAR dataset. The results identified that the AVODL-GRSM technique accurately recognizes six activities. For instance, on 70% of TRP, the AVODL-GRSM technique obtains average $accu_y$ of 99.43%, $prec_n$ of 98.32%, $reca_l$ of 98.24%, F_{score} of 98.27%, and MCC of 97.93%. In addition, on 30% of TSP, the AVODL-GRSM technique gains average $accu_y$ of 99.64%, $prec_n$ of 98.94%, $reca_l$ of 98.93%, F_{score} of 98.93%, and MCC of 98.72%.

Figure 4 inspects the accuracy of the AVODL-GRSM method in the training and validation on the UCI HAR database. The result notifies that the AVODL-GRSM technique has greater accuracy values over higher epochs. Also, the greater validation accuracy over training accuracy depicted that the AVODL-GRSM method learns productively on the UCI HAR database.

The loss analysis of the AVODL-GRSM method in the training and validation is pointed out on UCI HAR database

Table 2: Gesture recognition outcome of the AVODL-GRSM system on the UCI HAR dataset.

UCI HAR dataset					
Class	**Accuracy**	**Precision**	**Recall**	**F-score**	**MCC**
Training phase (70%)					
Walking (C-1)	99.35	97.81	98.37	98.09	97.69
Walking upstairs (C-2)	99.46	98.07	98.34	98.20	97.88
Walking downstairs (C-3)	99.50	98.95	97.30	98.12	97.83
Sitting (C-4)	99.51	98.95	98.24	98.59	98.30
Standing (C-5)	99.53	99.62	97.85	98.73	98.45
Laying/sleeping (C-6)	99.21	96.51	99.33	97.90	97.43
Average	99.43	98.32	98.24	98.27	97.93
Testing phase (30%)					
Walking (C-1)	99.68	98.80	99.20	99.00	98.81
Walking upstairs (C-2)	99.68	98.49	99.35	98.92	98.73
Walking downstairs (C-3)	99.68	99.32	98.42	98.86	98.68
Sitting (C-4)	99.64	98.68	99.24	98.96	98.75
Standing (C-5)	99.68	99.82	98.38	99.09	98.90
Laying/sleeping (C-6)	99.51	98.52	99.01	98.77	98.47
Average	99.64	98.94	98.93	98.93	98.72

Abbreviation: AVODL-GRSM, African Vulture Optimization with Deep Learning-based Gesture Recognition for Visually Impaired People on Sensory Modality Data.
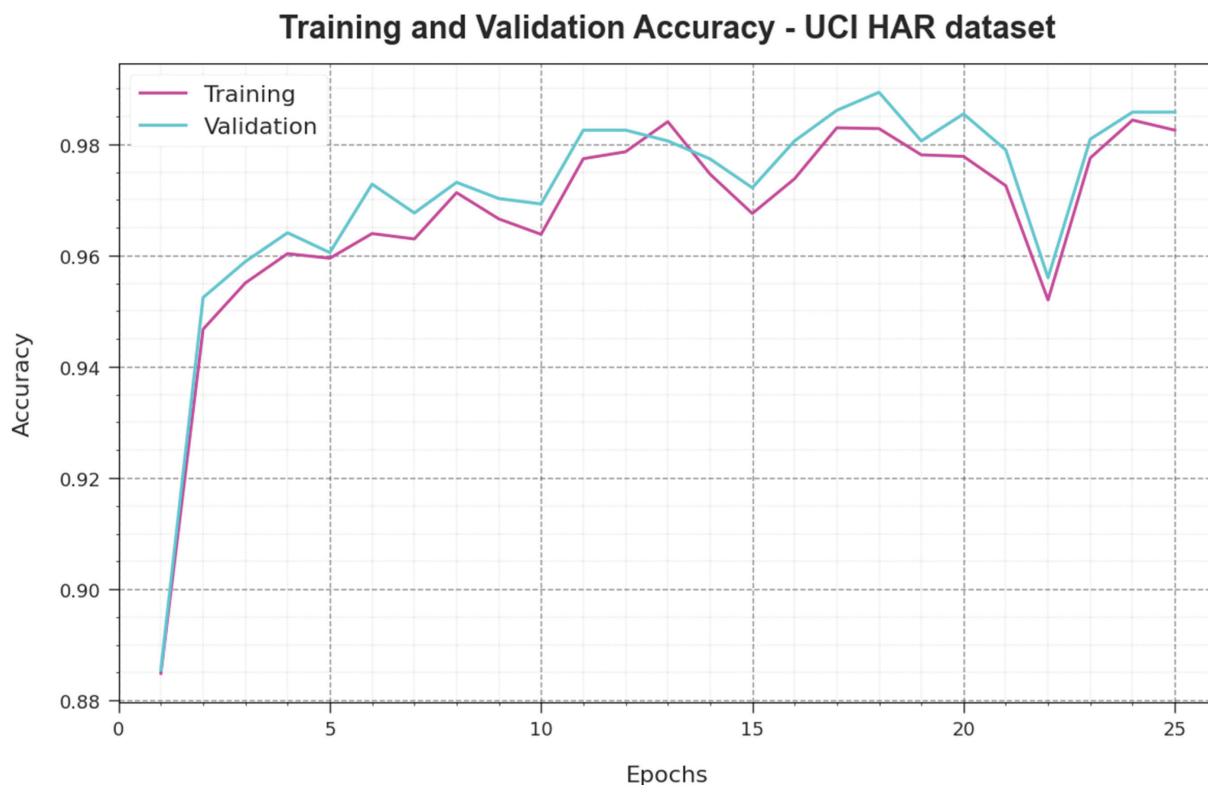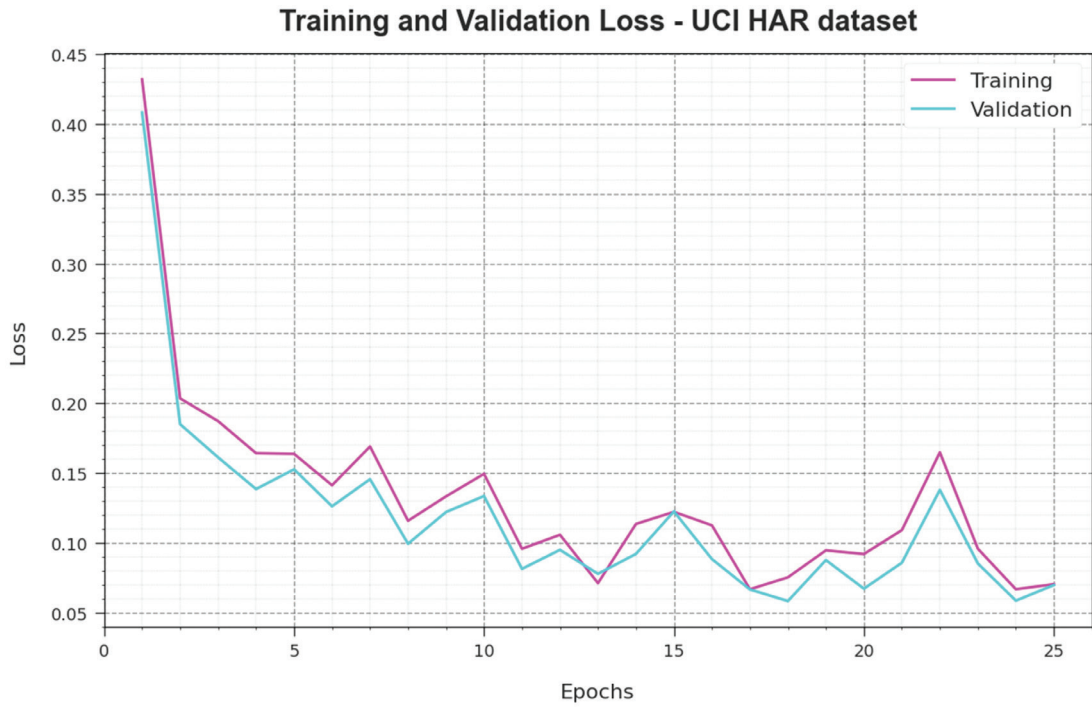

Figure 4: Accuracy curve of the AVODL-GRSM system on the UCI HAR dataset. Abbreviation: AVODL-GRSM, African Vulture Optimization with Deep Learning-based Gesture Recognition for Visually Impaired People on Sensory Modality Data.

Figure 5: Loss curve of the AVODL-GRSM system on the UCI HAR dataset. Abbreviation: AVODL-GRSM, African Vulture Optimization with Deep Learning-based Gesture Recognition for Visually Impaired People on Sensory Modality Data.
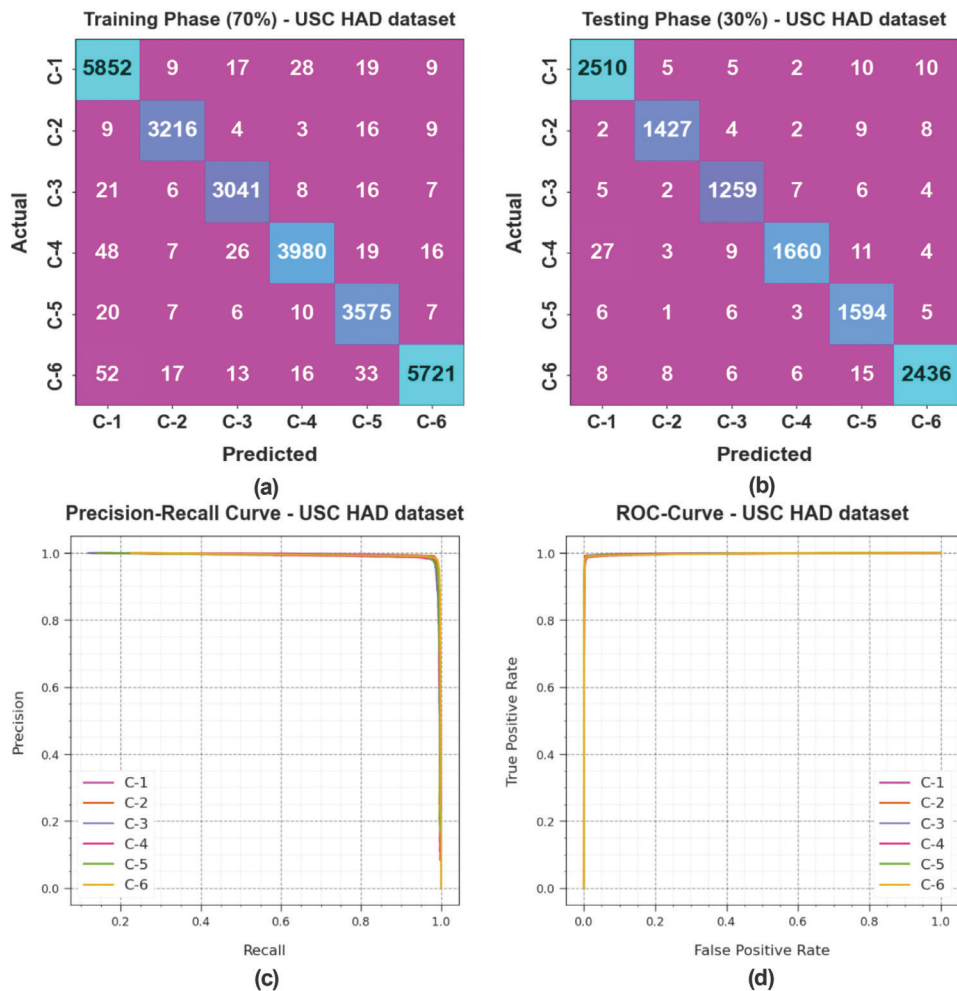

Figure 6: Classifier result on the USC HAD dataset. (a and b) Confusion matrices, (c) PR curve, and (d) ROC curve.

Table 3: Gesture recognition outcome of AVODL-GRSM system on USC HAD dataset.

USC HAD dataset					
Class	Accuracy	Precision	Recall	F-score	MCC
Training phase (70%)					
Walking (C-1)	99.10	97.50	98.62	98.06	97.48
Walking upstairs (C-2)	99.66	98.59	98.74	98.67	98.47
Walking downstairs (C-3)	99.52	97.88	98.13	98.00	97.73
Sitting (C-4)	99.30	98.39	97.17	97.78	97.36
Standing (C-5)	99.41	97.20	98.62	97.90	97.56
Laying/sleeping (C-6)	99.31	99.17	97.76	98.46	98.02
Average	99.38	98.12	98.17	98.14	97.77
Testing phase (30%)					
Walking (C-1)	99.28	98.12	98.74	98.43	97.96
Walking upstairs (C-2)	99.60	98.69	98.28	98.48	98.25
Walking downstairs (C-3)	99.51	97.67	98.13	97.90	97.63
Sitting (C-4)	99.33	98.81	96.85	97.82	97.43
Standing (C-5)	99.35	96.90	98.70	97.79	97.42
Laying/sleeping (C-6)	99.33	98.74	98.27	98.50	98.07
Average	99.40	98.16	98.16	98.15	97.79

Abbreviation: AVODL-GRSM, African Vulture Optimization with Deep Learning based Gesture Recognition for Visually Impaired People on Sensory Modality Data.

in Figure 5. The results indicate that the AVODL-GRSM method has adjacent values of training and validation loss. The AVODL-GRSM technique learns productively on the UCI HAR database.

Figure 6 shows the classifier results of the AVODL-GRSM method on the USC HAD dataset. Figure 6a and b illustrates the confusion matrix rendered by the AVODL-GRSM approach on 70:30 of TRP/TSP. The result specifies that the AVODL-GRSM approach has identified and classified all six class labels accurately. Likewise, Figure 6c shows the PR analysis of the AVODL-GRSM method. The result stated that the AVODL-GRSM algorithm has acquired higher PR performance under six classes. Eventually, Figure 6d presents the ROC investigation of the AVODL-GRSM model. The figure portrays that the AVODL-GRSM approach has productive outcomes with maximum ROC values under six class labels.
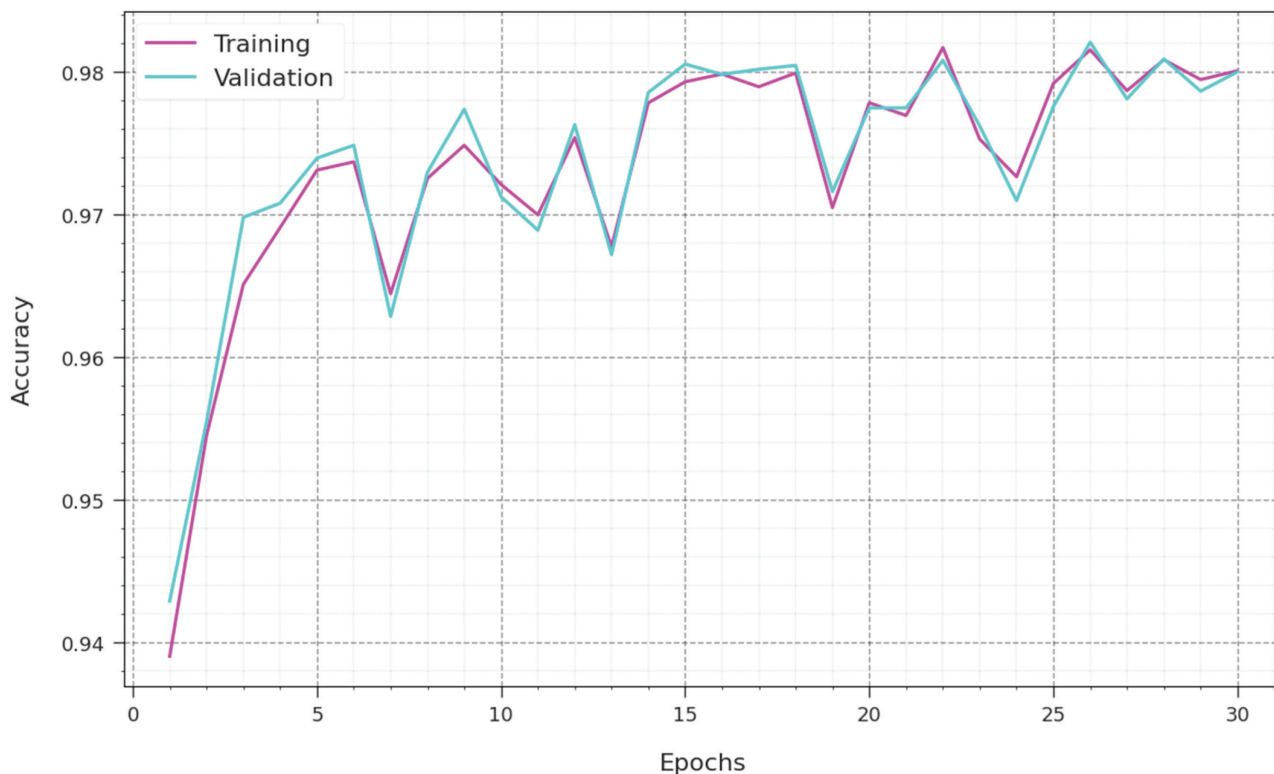

Figure 7: Accuracy curve of the AVODL-GRSM system on the USC HAD dataset. Abbreviation: AVODL-GRSM, African Vulture Optimization with Deep Learning-based Gesture Recognition for Visually Impaired People on Sensory Modality Data.
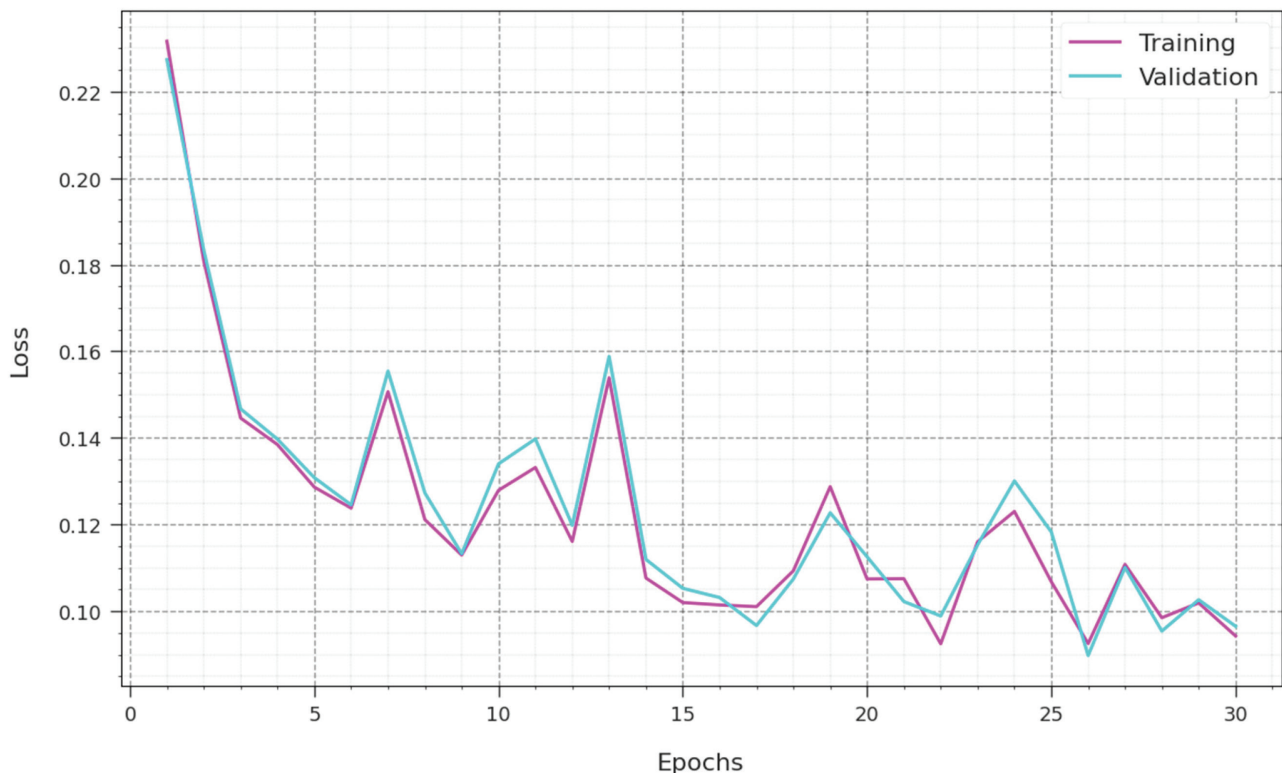
Figure 8: Loss curve of the AVODL-GRSM system on the USC HAD dataset. Abbreviation: AVODL-GRSM, African Vulture Optimization with Deep Learning-based Gesture Recognition for Visually Impaired People on Sensory Modality Data.

In Table 3, a brief recognition result of the AVODL-GRSM technique is clearly portrayed on the USC HAD dataset. The results identified that the AVODL-GRSM technique accurately recognizes six activities. For instance, on 70% of TRP, the AVODL-GRSM technique obtains average $accu_y$ of 99.43%, $prec_n$ of 98.32%, $reca_l$ of 98.24%, F_{score} of 98.27%, and MCC of 97.93%. In addition, on 30% of TSP, the AVODL-GRSM technique obtains average $accu_y$ of 99.64%, $prec_n$ of 98.94%, $reca_l$ of 98.93%, F_{score} of 98.93%, and MCC of 98.72%.

Figure 7 examines the accuracy of the AVODL-GRSM approach in training and validation on the USC HAD database. The figure notifies that the AVODL-GRSM technique has greater accuracy values over higher epochs. Moreover, the greater validation accuracy over training accuracy portrayed that the AVODL-GRSM approach learns efficiently on the USC HAD database.

The loss analysis of the AVODL-GRSM method in training and validation is shown on the USC HAD database in Figure 8. The results indicate that the AVODL-GRSM approach has adjacent values of training and validation loss. The AVODL-GRSM technique learns efficiently on the USC HAD database.

In Table 4 and Figure 9, the comparative outcomes of the AVODL-GRSM technique on two datasets are provided (Tahir et al., 2023). The results exhibited that the AVODL-GRSM technique reaches effective recognition results on both datasets. For instance, on the UCI HAR dataset, the AVODL-GRSM technique provides increasing $accu_y$ of 99.64% while the existing MWHODL-SHAR, convolutional

Table 4: Comparative outcome of AVODL-GRSM approach with other systems on two datasets.

Accuracy (%)		
Methods	UCI HAR dataset	USC HAD dataset
AVODL-GRSM	99.64	99.40
MWHODL-SHAR	99.09	99.03
CNN-RF	96.27	97.84
Residual network	95.45	95.86
Deep CNN	94.20	94.06
CAE	97.94	94.73
HARSI	95.86	95.76
LSTM	97.38	96.74

Abbreviation: AVODL-GRSM, African Vulture Optimization with Deep Learning based Gesture Recognition for Visually Impaired People on Sensory Modality Data.

neural network-random forest (CNN-RF), residual network, Deep CNN, CAE, human activity recognition on signal images (HARSI), and LSTM models provide decreasing $accu_y$ of 99.09, 96.27, 95.45, 94.20, 97.94, 95.86, and 97.38%, respectively. Also, on the USC HAD dataset, the AVODL-GRSM technique provides increasing $accu_y$ of 99.40% while the existing MWHODL-SHAR, CNN-RF, Residual network, Deep CNN, CAE, HARSI, and LSTM methods provide decreasing $accu_y$ of 99.03, 97.84, 95.86, 94.06, 94.73, 95.76, and 96.74%, correspondingly.

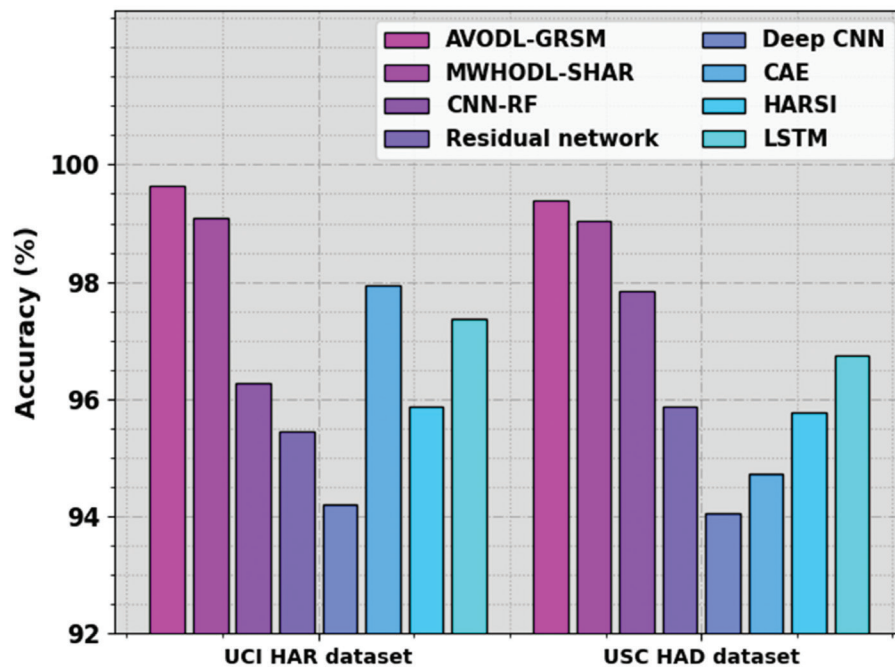These outcomes ensured the better performance of the AVODL-GRSM technique over other current methods.

Figure 9: Comparative outcome of the AVODL-GRSM approach with other systems on two datasets. Abbreviation: AVODL-GRSM, African Vulture Optimization with Deep Learning-based Gesture Recognition for Visually Impaired People on Sensory Modality Data.

CONCLUSION

This study has focused on the development of an automated gesture recognition tool named the AVODL-GRSMD technique for visually impaired people on sensory modality data technique. The AVODL-GRSMD technique exploits the DL model with hyperparameter tuning strategy for effectual and accurate gesture detection and classification process. The AVODL-GRSMD technique follows three major processes, namely data preprocessing, MHA-BGRU recognition, and AVO-based hyperparameter tuning. In this work, the hyperparameter optimization of the MHA-BGRU method is performed using the AVO algorithm. A series of simulation analyses were effectuated to demonstrate the improved performance of the AVODL-GRSMD technique. The experimental values demonstrate the better recognition rate of the AVODL-GRSMD technique compared to that of the state-of-the-art models.

ACKNOWLEDGMENTS

CONFLICTS OF INTEREST

The authors declare no conflicts of interest in association with the present study.

DATA AVAILABILITY STATEMENT

Data sharing not applicable to this article as no datasets were generated during the current study.

REFERENCES

Adeel M.I., Asad M.A., Zeeshan M.R., Amna M., Aslam M. and Martinez-Enriquez A.M. (2022). Gesture based confidence assessment system for visually impaired people using deep learning. In: *Advances in Information and Communication: Proceedings of the 2022 Future of Information and Communication Conference (FICC)*, Springer International Publishing, Cham, Vol. 2; pp. 135-147.

Agarwal A. and Das A. (2023). Facial gesture recognition based real time gaming for physically impairment. In: *Artificial Intelligence: First International Symposium, ISAI 2022*, Haldia, India, February 17-22, 2022, Springer Nature Switzerland, Cham, Revised Selected Papers; pp. 256-264.

Al-Hammadi M., Muhammad G., Abdul W., Alsulaiman M., Bencherif M.A. and Mekhtiche M.A. (2020). Hand gesture recognition for sign language using 3DCNN. *IEEE Access*, 8, 79491-79509.

Bao K., Bi J., Gao M., Sun Y., Zhang X. and Zhang W. (2022). An improved ship trajectory prediction based on AIS data using MHA-BiGRU. *J. Mar. Sci. Eng.*, 10(6), 804.

Can C., Kaya Y. and Kılıç F. (2021). A deep convolutional neural network model for hand gesture recognition in 2D near-infrared images. *Biomed. Phys. Eng. Express*, 7(5), 055005.

Deepa S., Umamageswari A. and Menaka S. (2023). A novel hand gesture recognition for aphonic people using convolutional neural network. In: *Computer Vision and Machine Intelligence Paradigms for SDGs:*

Select Proceedings of ICRTAC-CVMIP 2021, Springer Nature Singapore, Singapore; pp. 235-243,

Gangrade J. and Bharti J. (2023). Vision-based hand gesture recognition for Indian sign language using convolution neural network. *IETE J. Res.*, 69(2), 723-732.

Gorobets V., Merkle C. and Kunz, A. (2022). Pointing, pairing and grouping gesture recognition in virtual reality. In: *Computers Helping People with Special Needs: 18th International Conference, ICCHP-AAATE 2022*, Lecco, Italy, July 11–15, 2022, Springer International Publishing, Cham, Proceedings, Part I; pp. 313-320.

Kraljević L., Russo M., Pauković M. and Šarić M. (2020). A dynamic gesture recognition interface for smart home control based on Croatian sign language. *Appl. Sci.*, 10(7), 2300.

Lahiani H. and Neji M. (2018). Hand gesture recognition method based on HOG-LBP features for mobile devices. *Procedia Comput. Sci.*, 126, 254-263.

Li J., Li C., Han J., Shi Y., Bian G. and Zhou S. (2022). Robust hand gesture recognition using HOG-9ULBP features and SVM model. *Electronics*, 11(7), 988.

Liu Q., Kosarirad H., Meisami S., Alnowibet K.A. and Hoshyar, A.N. (2023). An optimal scheduling method in IoT-fog-cloud network using combination of Aquila optimizer and African vultures optimization. *Processes*, 11(4), 1162.

Padmavathi R. (2021). Expressive and Deployable Hand Gesture Recognition for Sign Way of Communication for Visually Impaired People.

Pandey S. (2023). Automated gesture recognition and speech conversion tool for speech impaired. In: *Proceedings of Third International Conference on Advances in Computer Engineering and Communication Systems: ICACECS 2022*, Springer Nature Singapore, Singapore; pp. 467-476.

Ryumin D., Ivanko D. and Ryumina E. (2023). Audio-visual speech and gesture recognition by sensors of mobile devices. *Sensors*, 23(4), 2284.

Sahana T., Basu S., Nasipuri M. and Mollah A.F. (2022). MRCS: multi-radii circular signature based feature descriptor for hand gesture recognition. *Multimed. Tools Appl.*, 81(6), 8539-8560.

Tahir B.S., Ageed Z.S., Hasan S.S. and Zeebaree S.R.M. (2023). Modified wild horse optimization with deep learning enabled symmetric human activity recognition model. *Comput. Mater. Contin.*, 75(2), 4009-4024.

UCI HAR Dataset. https://www.kaggle.com/competitions/uci-har/data?select=UCI+HAR+Dataset+for+Kaggle.

USC HAD Dataset. https://sipi.usc.edu/had/.

Varsha M. and Nair C.S. (2021). Indian sign language gesture recognition using deep convolutional neural network. In: *2021 8th International Conference on Smart Computing and Communications (ICSCC)*, IEEE; pp. 193-197.

Zhang J. and Zeng X. (2022). Multi-touch gesture recognition of Braille input based on Petri Net and RBF Net. *Multimed. Tools Appl.*, 81(14), 19395-19413.