Sand Cat Swarm Optimizer with Deep Wavelet Autoencoder-based Sign Language Recognition for Hearing- and Speech-impaired Persons

Mashael M. Asiri[1,2,*], Abdelwahed Motwakel[3] and Suhanda Drar[4]

[1]Department of Computer Science, College of Science and Arts at Mahayil, King Khalid University, Abha, Saudi Arabia
[2]King Salman Center for Disability Research, Riyadh, Saudi Arabia
[3]Department of Information Systems, College of Business Administration in Hawtat Bani Tamim, Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia
[4]Department of Computer and Self Development, Preparatory Year Deanship, Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia

Correspondence to:
Mashael M. Asiri*, e-mail: abusharara@kku.edu.sa

ABSTRACT

Sign language is commonly used to interact with people who have speech and hearing disorders. Sign language was exploited for interacting with people having developmental impairments who have some or no communication skills. Communication using Sign language has become a fruitful means of interaction for speech- and hearing-impaired people. The hand gesture recognition technique is useful for dumb and deaf people by using convolutional neural networks (CNNs) and human–computer interface for recognizing the static indication of sign language. Therefore, this study presents a new Sand Cat Swarm Optimizer with Deep Wavelet Autoencoder-based Intelligent Sign Language Recognition (SCSO-DWAESLR) technique for hearing- and speech-impaired persons. In the presented SCSO-DWAESLR technique, computer vision and CNN concepts are utilized for identifying sign languages to aid the interaction of hearing- and speech-impaired persons. The SCSO-DWAESLR method makes use of the Inception v3 model for the feature map generation process. In addition, the DWAE classifier is utilized for the recognition and classification of different kinds of signs posed by hearing- and speech-impaired persons. Finally, the hyperparameters related to the DWAE classifier are optimally chosen by using the SCSO algorithm. For exhibiting the effectual recognition outcomes of the SCSO-DWAESLR technique, a detailed experimental analysis was performed. The comparative outcome highlights the superior recognition performance of the SCSO-DWAESLR method over existing techniques under several evaluation metrics.

KEYWORDS

deep learning, sign language recognition, hearing and speech impairment, computer vision, convolutional neural network

INTRODUCTION

Sign language was introduced for hearing- and speech-impaired individuals to interact efficiently (Bora et al., 2023). A language refers to a system of interaction containing a set of written symbols and sounds a person utilizes. However, speech- and hearing-impaired persons are not able to use that language to interact; instead, they utilize sign language. A standardized global form of sign language does not exist and consequently, sign languages from various regions or nations are not mutually intelligible (Katoch et al., 2022). Deaf individuals are dependent on sign language as a means of interaction in day-to-day life (Novopoltsev et al., 2023). Hearing impairment can be categorized as severe, mild, moderate, or profound based on the severity of the deafness. People with profound or severe hearing disorders cannot attend to

others and therefore have transmission complexity. This poor transmission could affect the mental health of deaf persons, including solitude dissatisfaction (Mannan et al., 2022). The deaf society communicates with the help of gesture-related language, termed sign language. Late detection language is known as sign language. Hearing-impaired persons make use of sign language motions to connect (Das et al., 2023). The hearing society does not detect some gestures that constitute a transmission between a hearing and deaf individual. Sign language refers to a system of expression using signs. Mostly, hand gestures and facial expressions were utilized for transmission. Sign language is also a form of natural language (Rwelli et al., 2022). Sign language has its own set of vocabulary and alphabets. In deaf cultures, sign languages

have served a crucial role in linking people related to cultural and ethnic similarities. Sign language relies on facial expressions, hand gestures, and body movements (Duy Khuat et al., 2021). In sign language communication, lip gestures, facial expressions, hands, and eyes are used to send data.

Currently, Sign Language Recognition (SLR) has grabbed the attention of many. Currently, DL and ML have attained prominent development in the fields of recognition and classification (Aarthi et al., 2023). The efficacy of an SLR mechanism relies upon how fast and accurately it tracks the features and orientations of the hand. There are many algorithms for machine learning and FS (feature extraction). In the area of DL, Convolutional Neural Networks (CNNs) are the optimal methods leveraged in feature extraction (Herath and Ishanka, 2022). With the development and discovery of CNN approaches, it has achieved a new height of success. Implementing CNN methods in the automated detection of sign language is preferred among authors. A SLR method could break down the transmission barrier between hearing-impaired persons and normal people by building a transmission bridge (Grover et al., 2021). Allowing persons with hearing disorders to contribute to political, social, and economic actions that could assist them become useful assets to society.

This study introduces a new Sand Cat Swarm Optimizer with Deep Wavelet Autoencoder-based Intelligent Sign Language Recognition (SCSO-DWAESLR) technique for hearing- and speech-impaired persons. In the presented SCSO-DWAESLR technique, computer vision (CV) and CNN concepts are utilized for identifying sign language to aid communication between hearing- and speech-impaired persons. The SCSO-DWAESLR method makes use of the Inception v3 model for the feature map generation process. In addition, the DWAE classifier is utilized for the recognition and classification of different kinds of signs posed by hearing- and speech-impaired persons. Finally, the hyperparameters related to the DWAE classifier are optimally chosen by the use of the SCSO technique. For exhibiting the effectual recognition outcomes of the SCSO-DWAESLR technique, a detailed experimental analysis was performed.

LITERATURE REVIEW

In Eunice and Hemanth (2023) a deep analysis of the prevailing approaches is presented to devise a powerful system, deliberating their merits and demerits. Moreover, depending on the quality outcomes from the video generation unit, the author evaluated the performance of these methods. The author explained the future scope of launching real-time SLP transmission approaches constructed with advanced DL structures for the speech- and hearing-disabled leads to imparting employment and education to those people. Abeje et al. (2022) introduced an Ethiopian sign language, a new SLR mechanism that translated to Amharic alphabets using Deep CNN and CV technology. The technique accepts sign language imageries as input and presents Amharic text as the desirable output.

Li and Meng (2022) developed a multi-view spatiotemporal continuous SLR system. The network has three parts. The presented method has the capability of extracting the RGBs spatial, temporal attributes and skeleton datasets; the second is a sign language encoder system that depends on a transformer that could study long-term dependency; the last is a CTC expanded as a Connectionist Temporal Classification decoder system that can be employed for forecasting the meaning of the continuous sign language. In Sharma et al. (2021), a wider analysis of different gesture recognition techniques, including CNN and ML methods, was tested for real-time precision. Three methods: a pretrained VGG16 with finetuning, VGG16 with transfer learning (TL) and a hierarchical NN are examined on a counter of trainable variables.

Xu et al. (2021) intended to investigate the use of tensor-train decomposition in S2VT methods for minimalizing the parameter. First, the author systematically study the effect of variables of tensor-train factorizations on the method outcome. Then the author adopted tensor-train decomposition in various layers of the S2VT method for establishing six tensor-train S2VT approaches for Chinese SLR. Haque et al. (2019) developed the SLR structure to recognize 26 gestures from the Bangla Sign Language with the use of PYTHON. The PCA has been employed to recognize images by abstracting the principal component, and the KNN method was utilized for the classification stage. The authors Sruthi and Lijiya (2019) modeled a signer-independent DL-oriented technique to construct an Indian Sign Language (ISL) static alphabet detection technique. The author's analysis presents approaches in SLR and applies a CNN framework for recognizing the ISL static alphabet from the binary silhouette of the signer hand region.

THE PROPOSED MODEL

In this study, a new SCOS-DWAESLR method was introduced for the detection and classification of sign languages for hearing- and speech-impaired persons. In the presented SCSO-DWAESLR technique, CV and CNN concepts are utilized for identifying sign languages to aid the interaction of hearing and speech-impaired persons. The SCSO-DWAESLR method comprises a three-stage process such as Inception v3 feature extraction, DWAE classification, and SCSO-based hyperparameter tuning. Figure 1 illustrates the overall flow of the SCOS-DWAESLR system.

Feature extraction using Inception v3

In this work, the Inception v3 model is exploited for feature extraction. The network of Inception v3 is a DL technique. An effort takes place in training the model utilizing a low-configured computer when there is no access to a PC (Jena et al., 2023). Thus, Inception v3 performs efficiently with TL and an important graph of the Inception v3 technique. The TensorFlow library can be utilized for retraining Inception's last layer to novel types. A TL algorithm is a data-obtaining model which utilizes the preceding layer parameters and eliminates the last layer, afterwards retraining the final layer.
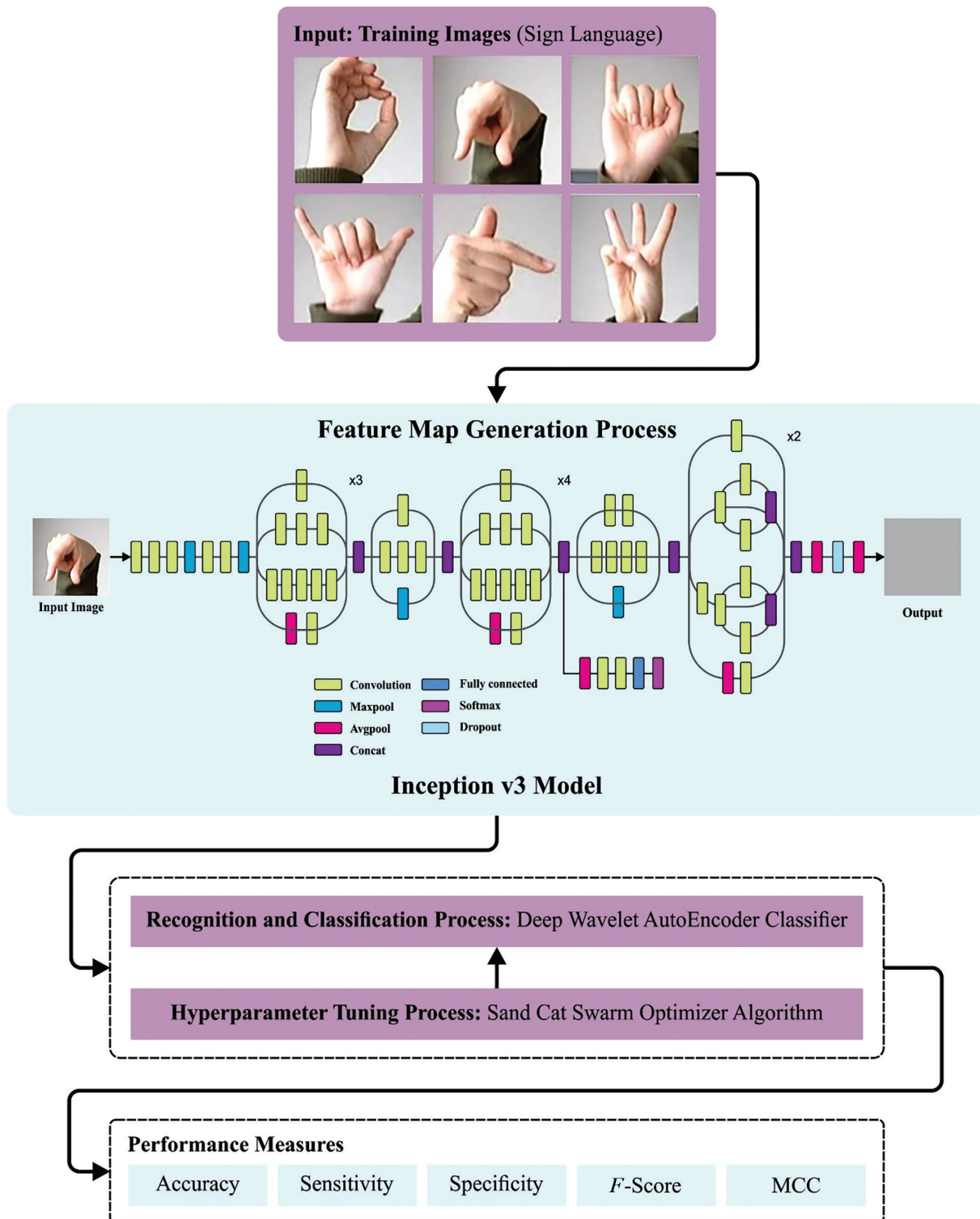
Figure 1: Overall flow of the SCOS-DWAESLR method.

The final layer of resultant nodes can be equivalent to the count of database categories. Thus, for the last classifier drives (0,1) the model loss is defined as follows:

$$Loss = \frac{\Sigma_i^N loss(i, class[i])}{\Sigma_i^N weight(class[i])} \qquad (1)$$

Training CNN frequently creates outcomes in over-fitting. Thus, this study executed pretraining CNNs to avoid this issue. The most popular CNN structures were utilized as DenseNet and ResNet. However, an optimum CNN technique which this study utilized is the Inception v3 technique, which is initialization by random weighted and finetuned on the database for feature extraction. The Inception v3 is a pretraining CNN technique which offers the optimum F1-score. This technique is the third type of inception family CNN technique considered by many enhancements. This technique gives an enhanced factorized convulsion that decreases the parameter count and retains network efficacy. This method utilizes regularized label smoothing.

Furthermore, an auxiliary classifier is utilized to help propagate label data and regularized them.

SLR using the DWAE model

The DWAE model is utilized for the detection of sign languages. AE is a kind of feedforward network where the input is similar to the output (Abd El Kader et al., 2021). Specifically, AE tries to recreate the output from the given representation and compresses the input vector into low-dimensional code. The AE includes three essential elements: the decoder, the encoder, and the code. Using the decoder, the output is reconstructed from that code.

The standard AE features a stronger inference capability, unsupervised feature learning capability, and robustness. The properties of Wavelet transform are time–frequency localization and focal features. Thus, it is important to combine typical AE and wavelet transform to resolve the real-time problem. This study presented a new kind of enhanced unsupervised NN named the "deep wavelet autoencoder" module. The wavelet AE exploited the wavelet function as an activation function, which defines different resolutions. Equation (2) specifies the decoding stage.

$$X = \xi(\hat{k}'Y + b') \tag{2}$$

In Equation (2), $\hat{X}$ denotes the outcome of the reconstructed vector, $\in$ refers to an error value added to the BP model, k shows the kernel vector, and b' represents the bias value.

The training process of the DWAE model is given as follows:

$$Training\ X = \xi(\hat{k}'Y + b')$$

Training instances $y = [y_1, y_2 \ldots y_n]A$ the output of the hidden unit is i.

$$g_i\ (out) = \varphi \frac{\left(\sum_{i=1}^{n} v_{il} y_l - e_i\right)}{b_i} \tag{3}$$

In Equation (3), φ characterizes the wavelet activation function.

$y_1(r = 1 \ldots 2 \ldots n)$ represents the lth dimension input of the training instance.

$V_{ij}(I = 1,2,3 \ldots g)$ denotes the weight connected between hidden units i and the input units.

b_i and e_i represent $V_{ij}(I = 1,2,3 \ldots g)$ sent the scale and shift factors of wavelet activation function for ith hidden units.

$$\varphi(a) = \cos(5a)\exp\left(\frac{a^2}{2}\right) \tag{4}$$

$$g_i(out) = \varphi_{b,e}(i) = \cos\left(5 \times \frac{\left(\sum_{i=1}^{n} v_{il} y_l - e_i\right)^2}{b_i}\right) \\ \times \exp\left(-\frac{1}{2}\left(\frac{\left(\sum_{l=1}^{n} v_{il} y_l - e_i\right)^2}{b_i}\right)\right) \tag{5}$$

The activation function of the output layer is designated as a sigmoid function like typical AE. T denotes the output of DWAE as follows:

$$\hat{y} = sigm\left(\sum_{i=1}^{q} v_{ri}\left(\cos\left(5 \times \frac{\left(\sum_{i=1}^{n} v_{il} y_l - e_i\right)}{b_i}\right)\right)\right) \\ \times \exp\left(\frac{-1}{2}\frac{\left(\sum_{l=1}^{n} v_{il} y_l - e_i\right)^2}{b_i}\right) \tag{6}$$

Where v_{ri} represents the weight connected between hidden r and $\hat{y}$ denotes i the reconstructed dimension output of the training instances.

Hyperparameter tuning using the SCSO algorithm

Lastly, the SCSO algorithm is employed for the hyperparameter selection of the DWAE model. SCSO technique is a new metaheuristic optimization approach. Sand cat (SC) lives in mountainous areas and barren deserts (Wang et al., 2023). The main source of food is Hares, Gerbils, insects, and snakes. In appearance, SCs look like domestic cats, but one dissimilarity is that their hearing can be highly sensitive and they could hear lower frequency noise lower than 2 kHz. Thus, they quickly use their special skills to track and attack the prey. The SC's predation is compared to the process of finding the optimum value, which is the main motivation of this proposed model.

The initialization process is performed randomly so that they can be uniformly distributed in the exploration region:

$$X_0 = lb + rand(0,1) \times (ub - lb) \tag{7}$$

In Equation (7), $rand$ shows the randomly generated value within [0, 1] and lb and ub denote the upper and lower boundaries.

The resultant initial matrix is given as follows:

$$Cat = \begin{bmatrix} x_{1,1} & x_{1,2} & \ldots & x_{1,M} \\ x_{2,1} & x_{2,2} & \ldots & x_{2,M} \\ \vdots & \vdots & \ldots & \vdots \\ x_{N,1} & x_{N,2} & \ldots & x_{N,M} \end{bmatrix} \tag{8}$$

In Equation (8), $x_{i,j}$ indicates the jth dimension at ith individuals, and there is an overall of N individuals and M parameters. However, the matrix of fitness function (FF) can be represented as follows:

$$Fitness = \begin{bmatrix} f(x_{1,1}; x_{1,2}; \ldots x_{1,M}) \\ f(x_{2,1}; x_{2,2}; \ldots x_{2,M}) \\ \vdots \\ f(x_{N,1}; x_{N,2}; \ldots x_{N,M}) \end{bmatrix} \tag{9}$$

The minimum value is found after comparing the fitness value, and the individual corresponding to it is the present optimum one.
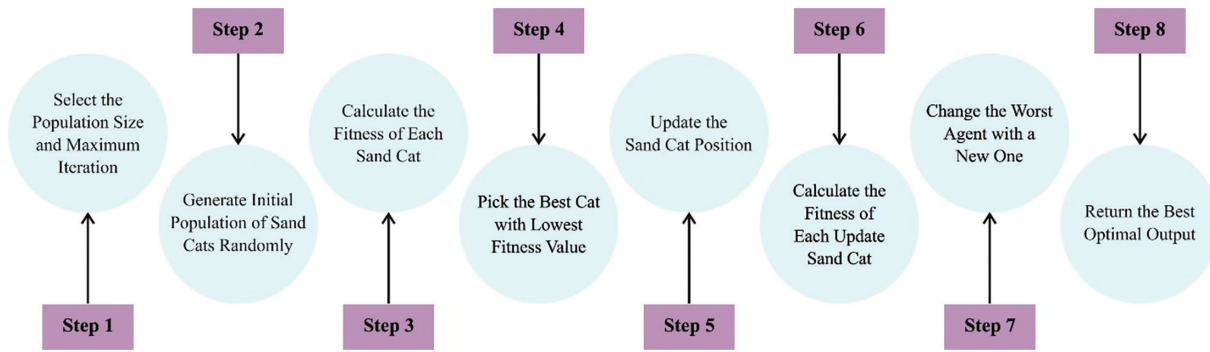
Figure 2: Steps involved in SCSO. Abbreviation: SCSO, Sand Cat Swarm Optimizer.

The SC finds prey by applying its sharp sense of hearing that could identify a low-frequency noise lower than 2 *kHz*:

$$S_e = S_M - \left(S_M \times \frac{t}{T} \right) \tag{10}$$

$$r_e = S_e \times rand\,(0,1) \tag{11}$$

$$X(t+1) = r_e \times (X_a(t) - rand\,(0,1) \times X(t)) \tag{12}$$

Here $X(t)$ indicates the immediate location of the SC, and $X(t)$ is any one of the populations. r_e symbolizes the sensitivity range of specific SC in the SC swarm. $S_M = 2$, S_e indicates the sensitivity range of the SCs, the value linearly declined from 2 to 0, t shows the immediate count of the iteration, and T denotes the maximum amount of iterations. Especially, if $S_e = 0$, $r_e = 0$, then the newest location of the SC will also be allocated to 0.

Moreover, R_e is put forward, and $R_e \in [0, 2]$ to ensure a steady state between the exploitation and exploration stages, and its value is shown below.

$$R_e = 2 \times S_e \times rand\,(0,1) - S_e \tag{13}$$

In the prior phase, the SCs attack the target as the search process progresses and it can be mathematically modeled as follows:

$$dist = \left| rand(0,1) \times X_{best}(t) - X(t) \right| \tag{14}$$

$$X(t+1) = X(t) - dist \times \cos(\theta) \times r_e \tag{15}$$

Where *dist* denotes the distance between the better and the existing individuals. θ shows the random angle from 0 to 360. Figure 2 represents the steps involved in the SCSO algorithm as explained in Algorithm 1 below.

The transformation of SCSO from the exploration to the exploitation phase is related to the variable R_e. If $|R_e|<1$, then the SC gets in the closet and captures the target which is in the exploitation stage; if $|R_e|>1$, it continues to search dissimilar spaces to search for the prey position which is in the exploration stage:

$$X(t+1) = \begin{cases} X_{best}(t) - dist \times \cos(\theta) \times r_e, \left| R_e \le 1 \right|; exploitation \\ r_e \times (X(t) - rand(0,1) \times X(t)), \left| R_e \ge 1 \right|; exploration \end{cases}$$
$$\tag{16}$$

Algorithm 1 Pseudocode of SCSO algorithm

Initialize the population
Evaluate the FF by using the objective function
Initialization process
While (t ≤ maximal iteration)
　For all the search agents
　　Attain a random angle by using Roulette Wheel Selection (0° ≤ α ≤ 360°)
　　If (abs(R) > 1)
　　　Upgrade the searching agent location by using Equation (12)
　　Else
　　　Upgrade the searching agenting location by using Equation (14)
　End
T = t + 1
End

The SCSO technique not only derives an FF to obtain superior classification accuracy and defines a positive integer to characterize the best outcome of the candidate solution. The decline of the classification error rate is assumed as an FF.

$$fitness(x_i) = ClassifierErrorRate(x_i)$$
$$= \frac{no.\,of\,misclassified\,instances}{Total\,no.of\,instances} \times 100 \tag{17}$$

Table 1: Details on database.

Labels	Class	Labels	Class
C1	A	C19	S
C2	B	C20	T
C3	C	C21	U
C4	D	C22	V
C5	E	C23	W
C6	F	C24	X
C7	G	C25	Y
C8	H	C26	Z
C9	I	C27	0
C10	J	C28	1
C11	K	C29	2
C12	L	C30	3
C13	M	C31	4
C14	N	C32	5
C15	O	C33	6
C16	P	C34	7
C17	Q	C35	8
C18	R	C36	9

RESULTS AND DISCUSSION

In this section, the SLR outcomes of the SCSO-DWAESLR method can be tested on the dataset including samples under 36 classes as shown in Table 1.

Table 2 reports the results of the SCSO-DWAESLR technique on 70% of TRP. The results illustrated that the SCSO-DWAESLR technique obtains improved recognition under 36 class labels.

For instance, with class 1, the SCSO-DWAESLR technique gains $accu_y$, $sens_y$, $spec_y$, F_{score}, and MCC of 98.73, 77.27, 99.31, 76.12, and 75.48%, respectively. Likewise, with class 10, the SCSO-DWAESLR method gains $accu_y$, $sens_y$, $spec_y$, F_{score}, and MCC of 98.97, 85, 99.43, 83.95,

and 83.42%, correspondingly. Similarly, with class 20, the SCSO-DWAESLR technique gains $accu_y$, $sens_y$, $spec_y$, F_{score}, and MCC of 99.56, 93.15, 99.75, 92.52, and 92.29% correspondingly. Finally, with class 36, the SCSO-DWAESLR method gains $accu_y$, $sens_y$, $spec_y$, F_{score}, and MCC of 99.05, 80.28, 99.59, 82.61, and 82.16%, correspondingly.

Table 3 reports the results of the SCSO-DWAESLR technique on 30% of TSP. The results illustrated that the SCSO-DWAESLR method obtains improved recognition under 36 class labels. For instance, with class 1, the SCSO-DWAESLR technique gains 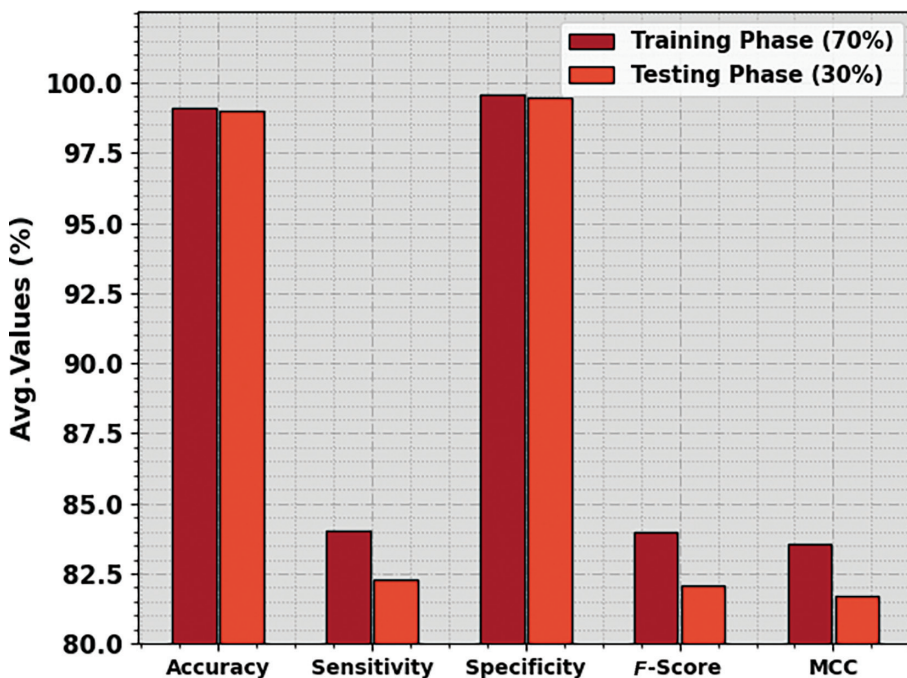$accu_y$, $sens_y$, $spec_y$, F_{score}, and MCC of 99.17, 88.24, 99.52, 86.96, and 86.54%, correspondingly. Likewise, with class 10, the SCSO-DWAESLR technique gains $accu_y$, $sens_y$, $spec_y$, F_{score}, and MCC of 99.63, 95, 99.72, 90.48,

Table 2: Classifier outcome of the SCSO-DWAESLR system on 70% of TRP.

Labels	$Accu_y$	$Sens_y$	$Spec_y$	F_{score}	MCC
Training phase (70%)					
1	98.73	77.27	99.31	76.12	75.48
2	99.09	80.88	99.59	82.71	82.26
3	99.21	82.86	99.67	85.29	84.93
4	99.29	91.55	99.51	87.84	87.55
5	99.17	81.82	99.63	83.72	83.32
6	99.29	88.00	99.63	88.00	87.63
7	99.09	84.06	99.51	83.45	82.99
8	99.01	79.10	99.55	80.92	80.43
9	99.01	77.27	99.59	80.31	79.87
10	98.97	85.00	99.43	83.95	83.42
11	98.97	83.33	99.43	82.19	81.67
12	99.05	81.25	99.51	81.25	80.76
13	99.17	85.25	99.51	83.20	82.80
14	98.89	90.00	99.14	81.82	81.61
15	99.29	88.89	99.59	87.67	87.31
16	99.29	86.30	99.67	87.50	87.14
17	99.21	83.78	99.67	86.11	85.74
18	99.09	82.86	99.55	83.45	82.99
19	98.93	73.91	99.63	79.07	78.72
20	99.56	93.15	99.75	92.52	92.29
21	99.25	86.89	99.55	84.80	84.44
22	99.09	86.30	99.47	84.56	84.11
23	99.13	88.06	99.43	84.29	83.92
24	99.13	85.48	99.47	82.81	82.41
25	99.05	84.72	99.47	83.56	83.08
26	99.13	91.18	99.35	84.93	84.69
27	99.21	81.16	99.71	84.85	84.53
28	99.01	80.28	99.55	82.01	81.52
29	99.13	83.12	99.63	85.33	84.92
30	99.29	88.57	99.59	87.32	86.97
31	99.33	85.33	99.75	88.28	87.98
32	99.01	78.57	99.59	81.48	81.03
33	98.93	80.56	99.47	81.12	80.57
34	99.29	89.33	99.59	88.16	87.80
35	98.93	78.57	99.51	80.29	79.76
36	99.05	80.28	99.59	82.61	82.16
Average	99.12	84.03	99.55	83.99	83.58

Abbreviation: SCSO-DWAESLR, Sand Cat Swarm Optimizer with Deep Wavelet autoencoder-based Intelligent Sign Language Recognition.

Table 3: Classifier outcome of the SCSO-DWAESLR system on 30% of TSP.

Labels	$Accu_y$	$Sens_y$	$Spec_y$	F_{score}	MCC
Testing phase (30%)					
1	99.17	88.24	99.52	86.96	86.54
2	98.80	78.12	99.43	79.37	78.76
3	99.17	86.67	99.52	85.25	84.83
4	99.26	93.10	99.43	87.10	86.91
5	99.07	79.41	99.71	84.38	84.07
6	98.80	68.00	99.53	72.34	71.88
7	99.17	77.42	99.81	84.21	84.13
8	98.61	78.79	99.24	77.61	76.90
9	99.07	84.85	99.52	84.85	84.37
10	99.63	95.00	99.72	90.48	90.39
11	99.17	85.71	99.52	84.21	83.80
12	98.98	83.33	99.52	84.51	83.99
13	98.43	69.23	99.52	76.06	75.64
14	99.07	86.67	99.43	83.87	83.44
15	98.52	75.00	99.14	72.41	71.70
16	98.06	62.96	98.96	61.82	60.83
17	99.26	84.62	99.62	84.62	84.24
18	99.07	83.33	99.52	83.33	82.86
19	99.07	87.10	99.43	84.38	83.94
20	99.44	88.89	99.72	88.89	88.60
21	98.89	79.49	99.62	83.78	83.34
22	98.89	74.07	99.53	76.92	76.41
23	99.26	87.88	99.62	87.88	87.50
24	99.17	86.84	99.62	88.00	87.58
25	98.98	82.14	99.43	80.70	80.19
26	99.44	81.25	100.00	89.66	89.88
27	99.44	90.32	99.71	90.32	90.04
28	98.43	79.31	98.95	73.02	72.45
29	98.80	78.26	99.24	73.47	73.00
30	98.98	96.67	99.05	84.06	84.31
31	99.17	92.00	99.34	83.64	83.58
32	98.43	66.67	99.33	70.18	69.47
33	98.98	85.71	99.33	81.36	80.94
34	99.35	80.00	99.81	85.11	84.96
35	98.98	83.33	99.43	81.97	81.45
36	99.44	82.76	99.90	88.89	88.86
Average	99.01	82.31	99.49	82.10	81.72

Abbreviation: SCSO-DWAESLR, Sand Cat Swarm Optimizer with Deep Wavelet autoencoder-based Intelligent Sign Language Recognition.

and 90.39%, correspondingly. Similarly, with class 20, the SCSO-DWAESLR method gains $accu_y$, $sens_y$, $spec_y$, F_{score}, and MCC of 99.44, 88.89, 99.72, 88.89, and 88.60%, correspondingly. Finally, with class 36, the SCSO-DWAESLR system gains $accu_y$, $sens_y$, $spec_y$, F_{score}, and MCC of 99.44, 82.76, 99.90, 88.89, and 88.86%, correspondingly.

Figure 3 reports the average recognition results of the SCSO-DWAESLR technique under 70% of TRP and 30% of TSP. The outcomes inferred that the SCSO-DWAESLR method obtains enhanced values under both classes. For instance, with 70% of TRP, the SCSO-DWAESLR technique gains average $accu_y$, $sens_y$, $spec_y$, F_{score}, and MCC of 99.12,

Figure 3: Average outcome of the SCSO-DWAESLR method on 70:30 of TRP/TSP. Abbreviation: SCSO-DWAESLR, Sand Cat Swarm Optimizer with Deep Wavelet autoencoder-based Intelligent Sign Language Recognition.
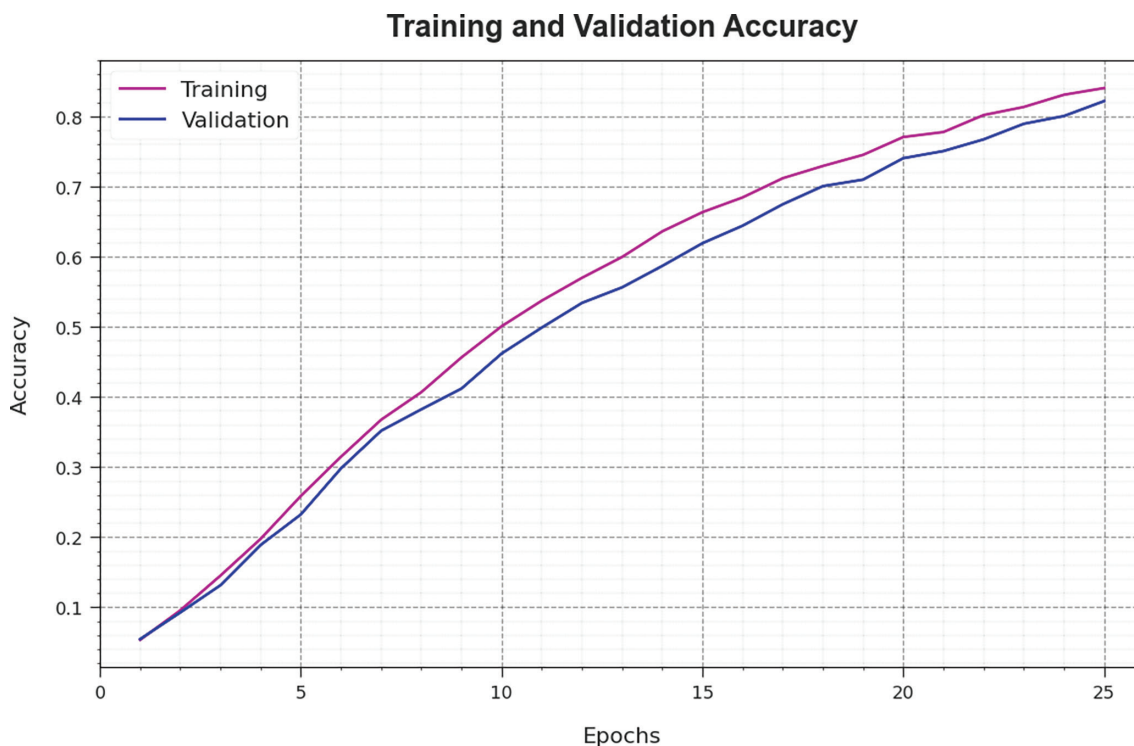

Figure 4: Accuracy curve of the SCSO-DWAESLR system. Abbreviation: SCSO-DWAESLR, Sand Cat Swarm Optimizer with Deep Wavelet autoencoder-based Intelligent Sign Language Recognition.

84.03, 99.55, 83.99, and 83.58%, respectively. Meanwhile, with 30% of TSP, the SCSO-DWAESLR technique gains average $accu_y$, $sens_y$, $spec_y$, F_{score}, and MCC of 99.01, 82.31, 99.49, 82.10, and 81.72%, correspondingly.

Figure 4 investigates the accuracy of the SCSO-DWAESLR method during the training and validation method on the test database. The figure indicates that the SCSO-DWAESLR technique obtains increasing accuracy

values over increasing epochs. Additionally, the increasing validation accuracy over training accuracy depicts that the SCSO-DWAESLR method learns efficiently on the test database.

The loss analysis of the SCSO-DWAESLR technique at the time of training and validation is illustrated on the test database in Figure 5. The results show that the SCSO-DWAESLR technique reaches closer values of training and
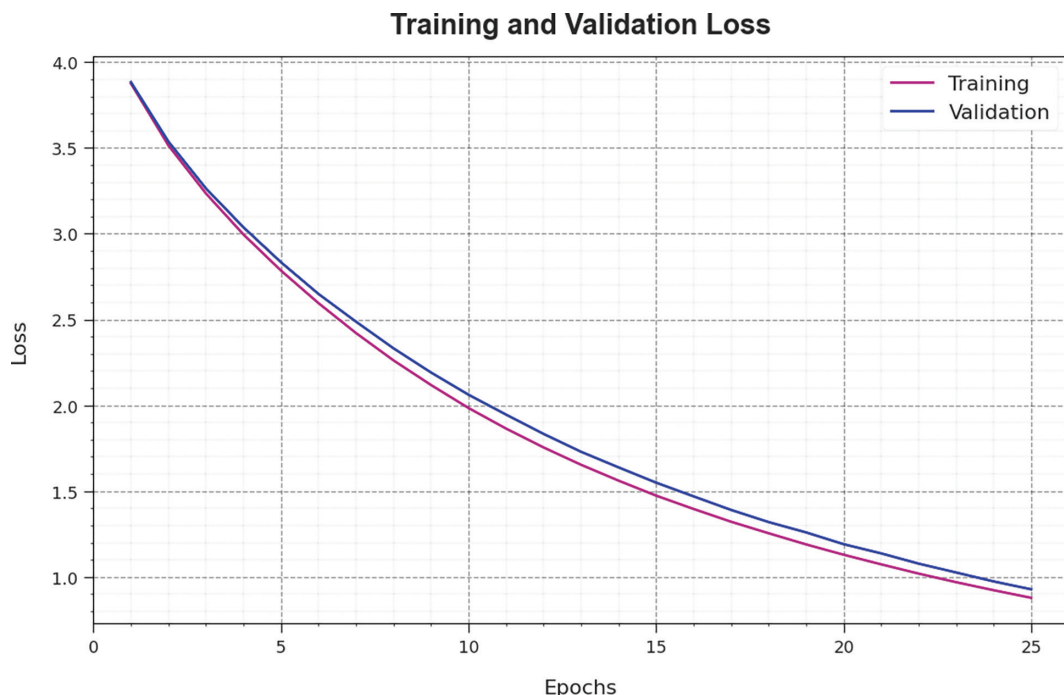

Figure 5: Loss curve of the SCSO-DWAESLR system. Abbreviation: SCSO-DWAESLR, Sand Cat Swarm Optimizer with Deep Wavelet autoencoder-based Intelligent Sign Language Recognition.
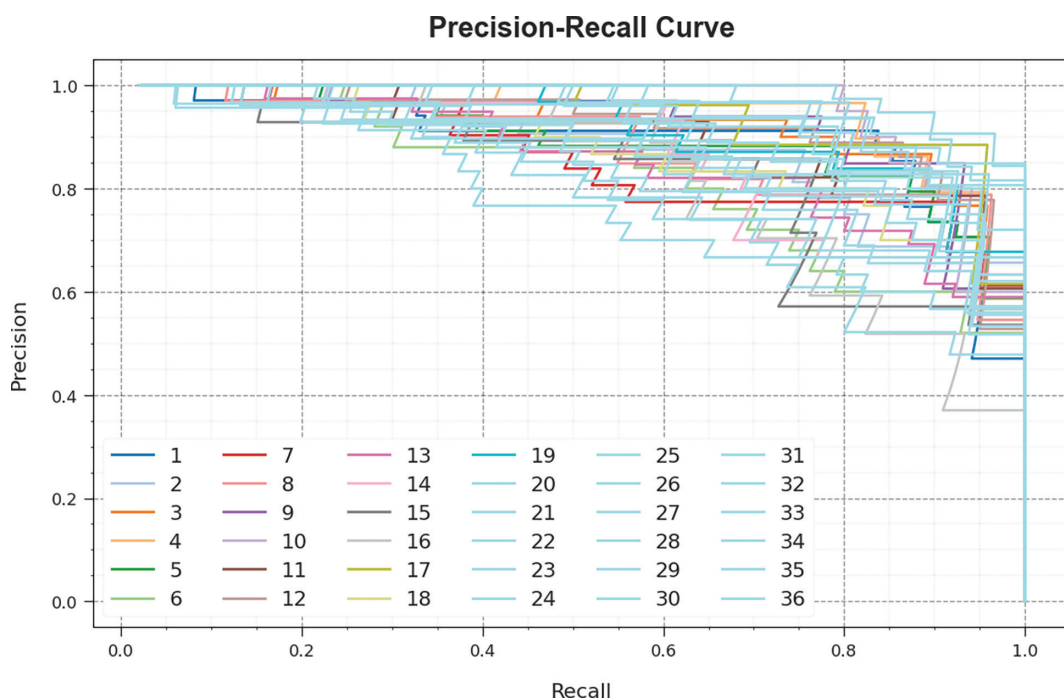

Figure 6: PR curve of the SCSO-DWAESLR system. Abbreviations: PR, precision-recall; SCSO-DWAESLR, Sand Cat Swarm Optimizer with Deep Wavelet autoencoder-based Intelligent Sign Language Recognition.
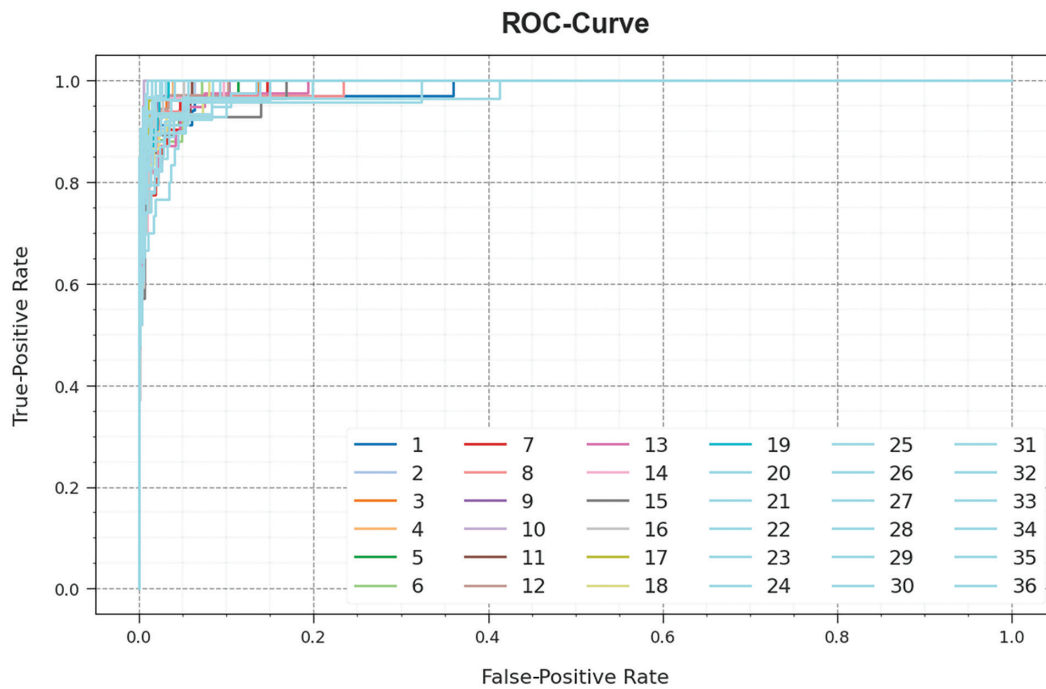
Figure 7: ROC curve of the SCSO-DWAESLR system. Abbreviation: SCSO-DWAESLR, Sand Cat Swarm Optimizer with Deep Wavelet autoencoder-based Intelligent Sign Language Recognition.

validation loss. The SCSO-DWAESLR method learns efficiently on a test database.

A brief precision-recall (PR) curve of the SCSO-DWAESLR technique is illustrated on the test database in Figure 6. The results stated that the SCSO-DWAESLR method results in increasing values of PR. In addition, it is noticeable that the SCSO-DWAESLR technique can reach higher PR values in all classes.

In Figure 7, a ROC study of the SCSO-DWAESLR system is revealed on the test database. The figure described that the SCSO-DWAESLR method resulted in improved ROC values. Besides, the SCSO-DWAESLR technique can extend enhanced ROC values on all classes.

To verify the improved results of the SCSO-DWAESLR technique, a widespread comparison study was performed and is shown in Table 4 and Figure 8 (Alnfiai, 2023). The

Table 4: Comparative outcome of the SCSO-DWAESLR system with other methods.

Methods	$Accu_y$	$Sens_y$	$Spec_y$
SCSO-DWAESLR	99.12	84.03	99.55
SSODL-ASLR	98.57	81.48	99.07
SVM model	93.28	79.45	96.07
Decision tree	89.92	76.59	93.15
KNN model	88.07	75.03	96.30
DNN model	91.52	77.72	93.23
LeNet model	92.14	74.64	92.90
MLP model	93.72	77.63	97.90

Abbreviation: SCSO-DWAESLR, Sand Cat Swarm Optimizer with Deep Wavelet autoencoder-based Intelligent Sign Language Recognition.

results demonstrate that the DT and KNN models accomplish the least performance. Next to that, the DNN, LeNet, MLP, and SVM models have managed to reach slightly improvised outcomes. Along with that, the SSODL-ASLR technique resulted in near-optimal performance with $accu_y$ of 98.57%, $sens_y$ of 81.48%, and $spec_y$ of 99.07%. Nevertheless, the SCSO-DWAESLR technique reaches higher performance with maximum $accu_y$ of 99.12%, $sens_y$ of 84.03%, and $spec_y$ of 99.55%. These outcomes show the best performance of the SCSO-DWAESLR method on the detection and classification of sign languages.

CONCLUSION

In this study, a new SCOS-DWAESLR technique is introduced for the identification and classification of sign languages for hearing- and speech-impaired persons. In the presented SCSO-DWAESLR technique, CV and CNN concepts are utilized for identifying sign languages to aid the interaction of hearing- and speech-impaired persons. The SCSO-DWAESLR method comprises a three-stage process such as Inception v3 feature extraction, DWAE classification, and SCSO-based hyperparameter tuning. Here, the hyperparameters related to the DWAE classifier are optimally chosen by using the SCSO technique. For exhibiting the effectual recognition outcomes of the SCSO-DWAESLR technique, a detailed experimental analysis was performed. The comparative outcome highlights the improved recognition performance of the SCSO-DWAESLR method over existing techniques under several evaluation metrics. In future, the computation complexity of the proposed model needs to be investigated.
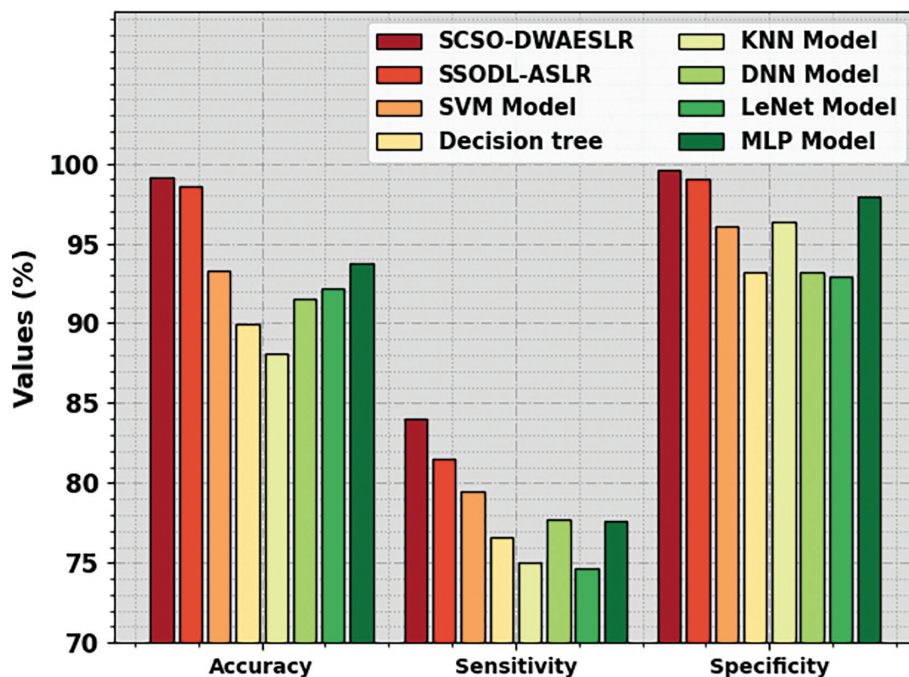
Figure 8: Comparative outcome of the SCSO-DWAESLR system with other approaches. Abbreviation: SCSO-DWAESLR, Sand Cat Swarm Optimizer with Deep Wavelet autoencoder-based Intelligent Sign Language Recognition.

FUNDING

CONFLICTS OF INTEREST

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

REFERENCES

Aarthi B., Charitharyan P., Thanmai R. and Kesana M. (2023). Implementation of dynamic gesture interpretation of sign language for impact on hearing and speech impairment. In: *Handbook of Research on Advanced Practical Approaches to Deepfake Detection and Applications*. IGI Global; pp. 257-274.

Abd El Kader I., Xu G., Shuai Z., Saminu S., Javaid I., Ahmad I.S., et al. (2021). Brain tumor detection and classification on MR images by a deep wavelet auto-encoder model. *Diagnostics*, 11(9), 1589.

Abeje B.T., Salau A.O., Mengistu A.D. and Tamiru N.K. (2022). Ethiopian sign language recognition using deep convolutional neural network. *Multimed. Tools Appl.*, 81(20), 29027-29043.

Alnfiai M.M. (2023). Deep learning-based sign language recognition for hearing and speaking impaired people. *Intell. Autom. Soft Comput.*, 36(2), 1653-1669.

Bora J., Dehingia S., Boruah A., Chetia A.A. and Gogoi D. (2023). Real-time assamese sign language recognition using MediaPipe and deep learning. *Procedia Comput. Sci.*, 218, 1384-1393.

Das S., Imtiaz M.S., Neom N.H., Siddique N. and Wang H. (2023). A hybrid approach for Bangla sign language recognition using deep transfer learning model with random forest classifier. *Expert Syst. Appl.*, 213, 118914.

Duy Khuat B., Thai Phung D., Thi Thu Pham H., Ngoc Bui A. and Tung Ngo S. (2021). Vietnamese sign language detection using MediaPipe. In: *2021 10th International Conference on Software and Computer Applications*. ACM, Kuala Lumpur, Malaysia, February 23-26, pp. 162-165.

Eunice R.J. and Hemanth D.J. (2023). Deep learning and sign language models based enhanced accessibility of e-governance services for speech and hearing-impaired. In: *Electronic Governance with Emerging Technologies: First International Conference, EGETC 2022*, Tampico, Mexico, 12-14 September 2022, Revised Selected Papers. Springer Nature Switzerland, Cham; pp. 12-24.

Grover Y., Aggarwal R., Sharma D. and Gupta P.K. (2021). Sign language translation systems for hearing/speech impaired people: a review. In: *2021 International Conference on Innovative Practices in Technology and Management (ICIPTM)*. IEEE, Noida, India, February 17-19, pp. 10-14.

Haque P., Das B. and Kaspy N.N. (2019). Two-handed Bangla sign language recognition using principal component analysis (PCA) and KNN algorithm. In: *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*. IEEE, Cox'sBazar, Bangladesh, February 7-9, pp. 1-4.

Herath R.J. and Ishanka P. (2022). An approach to Sri Lankan sign language recognition using deep learning with MediaPipe. In: *Digital Technologies and Applications: Proceedings of ICDTA'22*, Fez, Morocco, Volume 1. Springer International Publishing, Cham; pp. 449-459.

Jena R., Shanableh A., Al-Ruzouq R., Pradhan B., Gibril M.B.A., Khalil M.A., et al. (2023). Explainable artificial intelligence (XAI) model for earthquake spatial probability assessment in Arabian peninsula. *Remote Sens.*, 15(9), 2248.

Katoch S., Singh V. and Tiwary U.S. (2022). Indian sign language recognition system using SURF with SVM and CNN. *Array*, 14, 100141.

Li R. and Meng L. (2022). Multi-view spatial-temporal network for continuous sign language recognition. *arXiv preprint arXiv:2204.08747*.

Mannan A., Abbasi A., Javed A.R., Ahsan A., Gadekallu T.R. and Xin Q. (2022). Hypertuned deep convolutional neural network for sign language recognition. *Comput. Intell. Neurosci.*, 2022, 1450822.

Novopoltsev M., Verkhovtsev L., Murtazin R., Milevich D. and Zemtsova I. (2023). Fine-tuning of sign language recognition models: a technical report. *arXiv preprint arXiv:2302.07693*.

Rwelli R.E., Shahin O.R. and Taloba A.I. (2022). Gesture based Arabic sign language recognition for impaired people based on convolution neural network. *arXiv preprint arXiv:2203.05602*.

Sharma A., Sharma N., Saxena Y., Singh A. and Sadhya D. (2021). Benchmarking deep neural network approaches for Indian sign language recognition. *Neural Comput. Appl.*, 33, 6685-6696.

Sruthi C.J. and Lijiya A. (2019). Signet: a deep learning based indian sign language recognition system. In: *2019 International Conference on Communication and Signal Processing (ICCSP)*. IEEE, Chennai, India, April 4-6, pp. 0596-0600.

Wang X., Liu Q. and Zhang L. (2023). An adaptive sand cat swarm algorithm based on cauchy mutation and optimal neighborhood disturbance strategy. *Biomimetics*, 8(2), 191.

Xu B., Huang S. and Ye Z. (2021). Application of tensor train decomposition in S2VT model for sign language recognition. *IEEE Access*, 9, 35646-35653.