

# Deep Learning-Based Sign Language Recognition for Hearing and Speaking Impaired People

Mrim M. Alnfai\*

Department of Information Technology, College of Computers and Information Technology, Taif University, P.O. Box 11099, Taif, 21944, Saudi Arabia

\*Corresponding Author: Mrim M. Alnfai. Email: m.alnofee@tu.edu.sa

Received: 21 June 2022; Accepted: 04 August 2022

**Abstract:** Sign language is mainly utilized in communication with people who have hearing disabilities. Sign language is used to communicate with people having developmental impairments who have some or no interaction skills. The interaction via Sign language becomes a fruitful means of communication for hearing and speech impaired persons. A Hand gesture recognition system finds helpful for deaf and dumb people by making use of human computer interface (HCI) and convolutional neural networks (CNN) for identifying the static indications of Indian Sign Language (ISL). This study introduces a shark smell optimization with deep learning based automated sign language recognition (SSODL-ASLR) model for hearing and speaking impaired people. The presented SSODL-ASLR technique majorly concentrates on the recognition and classification of sign language provided by deaf and dumb people. The presented SSODL-ASLR model encompasses a two stage process namely sign language detection and sign language classification. In the first stage, the Mask Region based Convolution Neural Network (Mask RCNN) model is exploited for sign language recognition. Secondly, SSO algorithm with soft margin support vector machine (SM-SVM) model can be utilized for sign language classification. To assure the enhanced classification performance of the SSODL-ASLR model, a brief set of simulations was carried out. The extensive results portrayed the supremacy of the SSODL-ASLR model over other techniques.

**Keywords:** Sign language recognition; deep learning; shark smell optimization; mask rcnn model; disabled people

## 1 Introduction

Recently, the population of deaf-dumb victims has raised due to birth defects and other problems. A deaf and mute individual may not able to interact with ordinary people by relying certain kinds of transmission mechanisms [1]. The gesture displays certain physical actions of the hand which conveys a part of information. Gesture recognition was the analytical clarification of movement of a person via information processing mechanisms. Verbal communication offers the most effectual conversation platform for mute person for speaking with ordinary people [2]. Only some individual realizes the meaning of sign. Usually,



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

deaf person was deprived of usual contact with other typical individuals in the society. Human computer interface (HCI) was an exciting system amongst devices and individuals. This was an exciting research area which aims at the usage and formation of computer technology and particularly, combined communications amongst devices and humans. HCI structure was amazingly protracted and enhanced by the technical transition [3]. On the basis of fruitful usability, evolving technologies apply modern user lines like Speech recognition, Non-Touch, and Gesture. It is a complex and expensive technology to reach [4,5]. This newly applied technology was after combined as implying particular applications on the basis of cost-effectiveness and demand. For addressing the complexities, numerous authors were attempting for developing such interferences with regard to robustness, performance, and accessibility [6]. The optimum model could have several standard features like scalability, simplicity, flexibility, and precision. Nowadays, the human gesture turns to be an extensive HCI application, and the usage of human gestures, that fulfils all such standards, was rising quickly [7,8].

Sign language (SL) always considered the main way of verbal interaction amongst individuals who were both dumb and deaf [9]. Whereas interacting such people is very helpless and therefore were only depends on hand gestures. Visual signs and gestures were an energetic part of automated sign language (ASL) which offers deaf individuals reliable and easy communication [10]. It has well-defined code gestures in which every sign carries a specific meaning relating to communication. There were several methods to seek gestural data. But limiting to only major kinds there were 2 significant familiar kinds they are Vision-related and Sensor-related techniques [11]. The sensor-related technique gathers data from the glove produced by hand movement. In the vision-related technique, the image was considered by using cameras [12]. This technique indulges the image qualities like texture part and coloring which was necessary to the specific hand gesture.

In [13], a comparative analysis of different gesture recognition methods including convolutional neural network (CNN) and machine learning (ML) procedures has been deliberated and verified for realistic performance. A hierarchical neural network, pre-trained VGG16 with fine-tuning, and VGG16 with transfer learning were analyzed according to a trained parameter count. The model was trained on a self-developed datasets comprising images of Indian Sign Language (ISL) representation of each twenty-six English alphabet. In [14], the authors shows that a later fusion technique to multi-modality in sign language detection increases the entire capacity of algorithm when compared to singular approach of Leap Motion data and image classification. Using a larger synchronous dataset of eighteen BSL gestures gathered from different subjects, two deep neural networks (DNNs) were compared and benchmarked for deriving an optimal topology. The Vision model was carried out by a CNN and optimized artificial neural network (ANN), and the Leap Motion mechanism is carried out using an evolutionary search of ANN model.

Kumar et al. [15] developed the usage of graph matching (GM) to allow three dimensional motion capture for Indian sign language detection. The sign recognition and classification problems to interpret three dimensional motion signs are taken into account an adoptive GM (AGM) problems. But, the existing model to solve an AGM issue have two most important disadvantages. Firstly, spatial matching is implemented on a number of frames with a definite set of nodes. Then, temporal matching splits the whole three dimensional datasets into a definite set of pyramids. In [16], the authors conducted a complete systematic mapping of translation-assisting techniques to the provided sign language. The mapping has regarded the primary guideline for systematic review that is, pertains software engineering because it is necessary to take responsible for multidisciplinary fields of education, accessibility, human computer communication, and natural language processing. A continuous improvement of software tools named SYstematic Mapping and Parallel Loading Engine (SYMPLE) enabled the construction and querying of a base set of candidate studies. Parvez et al. [17] related the gap among conventional teaching and the technology-based methods that are employed to teach arithmetical concepts.

The participant was separated into developed mobile applications and conventional approaches (board and flash cards). The variance in the performance these groups is assessed by accompanying quizzes.

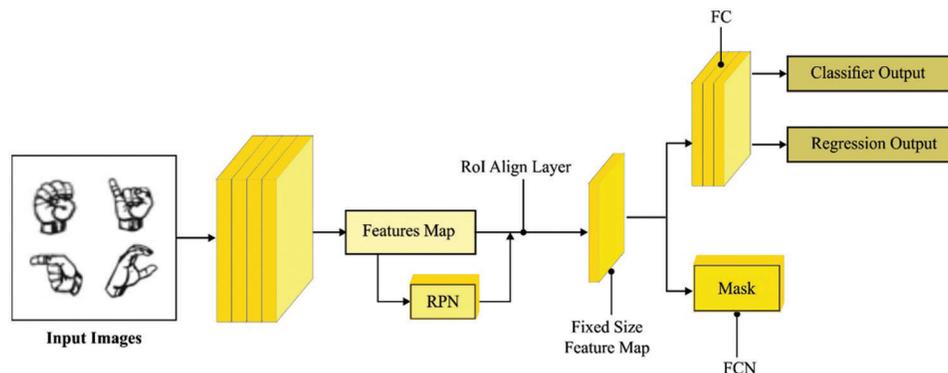
This study introduces a shark smell optimization with deep learning based automated sign language recognition (SSODL-ASLR) model for hearing and speaking impaired people. The presented SSODL-ASLR technique majorly concentrates on the recognition and classification of sign language provided by deaf and dumb people. The presented SSODL-ASLR model encompasses a two stage process namely sign language detection and sign language classification. At the first stage, the Mask Region based Convolution Neural Network (Mask RCNN) model is exploited for sign language recognition. Secondly, SSO algorithm with soft margin support vector machine (SM-SVM) model was utilized for sign language classification. To assure the enhanced classification performance of the SSODL-ASLR model, a brief set of simulations was carried out.

## 2 Design of SSODL-ASLR Model

In this study, a new SSODL-ASLR technique was introduced for the recognition of SL for hearing and speaking impaired people. The presented SSODL-ASLR technique majorly concentrates on the recognition and classification of sign language provided by deaf and dumb people. The presented SSODL-ASLR model encompasses a two stage process namely sign language detection and sign language classification.

### 2.1 Sign Language Detection: Mask RCNN Model

In the first stage, the Mask RCNN model is exploited for sign language recognition. Mask RCNN was a DNN mainly focused on solving instance segmentation issues in computer vision or ML [18]. In other words, it can separate different objects in an image or a video. The feature pyramid network (FPN) to object detection, a 1<sup>st</sup> block structure of Mask RCNN is accountable for extracting features. The regional proposal network (RPN), the 2<sup>nd</sup> part of Mask RCNN, and shares whole image convolutional features with detection network thus approximately assisting cost-free RPN. Once the extended Fast RCNN processes the Mask RCNN with added a branch to forecast an object mask from equivalent with the suggested branch to bound box recognition. The RPN is carried out from Mask RCNN instead of selective search and thus RPN shares the convolution features of entire map with detection network. It forecasts fused boundary location and objects score at each place, and it is fully convolution network (FCN). Based on the Mask RCNN, it presented an algorithm to improve the speed using side fusion feature pyramid network (SF-FPN) with Resnet-86 and enlighten the performance. In such cases, the dataset, FPN architecture, and RPN variable setting were improved. The improved method developed in such cases is appreciating the segmentation, recognition, and detection of target at the same time. Fig. 1 illustrates the infrastructure of Mask RCNN method.



**Figure 1:** Structure of mask RCNN

During the multitasking, loss function is carried out by trainable Mask R CNN with three segments namely classification locate regression loss of bounding box, loss of mask, and loss of bounding box as follows.

$$L = L_{cls} + L_{box} + L_{mask} \quad (1)$$

$$L_{cls} = -\log[p_i^* p_i + (1 - p_i^*)(1 - p_i)] \quad (2)$$

$$L_{box} = r(t_i - t_i^*) \quad (3)$$

$$L_{mask} = \text{Sigmoid}(Cls_k) \quad (4)$$

From the expression,  $p_i$  indicates the notable probability for ROI in classifier loss  $L_{cls}$  as well as  $p_i^*$  employed to ground truth as one that ROI was taken into account as zero or foreground.  $t_i$  represents the vector of accurate control to identify bounding boxes and  $t_i^*$  characterize the ground truth in location regression loss whereby  $r$  denotes the loss function to estimate the regression mistake. Each ROI detects the outcomes of  $K \times m^2$  dimension via mask branch and encoded  $K$  mask alongside resolution of  $m \times m$ . The loss of mask  $L_{mask}$  is regarded by Average Binary Cross-entropy Loss to carry out sigmoid function on each pixel from ROI. In class  $k(Cls_k)$ , mask loss was illustrated in above equation.

## 2.2 Sign Language Classification: SM-SVM Model

In this study, the SM-SVM model is utilized for sign language classification. SM-SVM aim is to expand the Maximal Separating Margin SVM (hard margin SVM), such that the hyperplane enables a noisy dataset to exist. In the event, a variable  $\xi_i$ , called Slack factor was presented to accountable for the number of violations of classifier [19].

$$Wx_i + b \geq 1 - \xi_i, \quad \forall y_i = +1, \quad (5)$$

$$Wx_i + b \leq -1 + \xi_i, \quad \forall y_i = -1;$$

That is,

$$y_i(Wx_i + b) \geq 1 - \xi_i, \quad i = 1, 2, \dots, n. \quad (6)$$

The Primal Problem is defined by

$$\min_{W,b} \frac{1}{2} \|W\|^2 + C \sum_{i=1}^n \xi_i \quad (7)$$

$$s.t. \quad y_i(Wx_i + b) \geq 1 - \xi_i,$$

$$\xi_i \geq 0, \quad i = 1, 2, \dots, n.$$

The Dual Problem Function of soft margin was equated by

$$\max_{\alpha} W(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i^T x_j$$

$$s.t. \quad \sum_{i=1}^n \alpha_i y_i = 0 \quad (8)$$

$$0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, n.$$

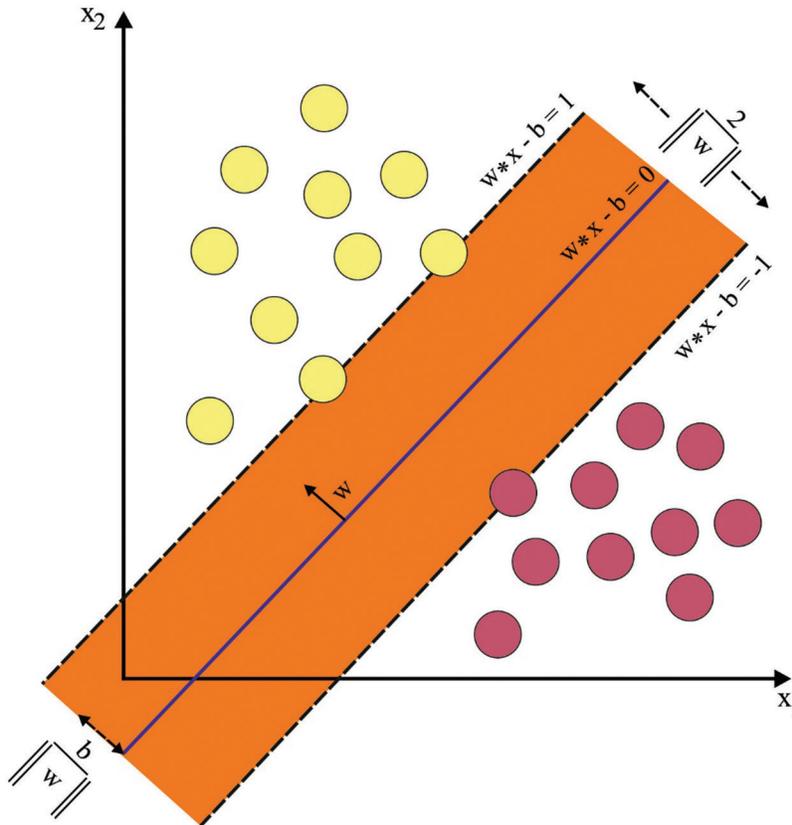
KKT complementary conditions in inseparable instance are represented as follows

$$\alpha_i [y_i(Wx_i + b) - 1 + \xi_i] = 0, \quad i = 1, 2, \dots, n, \tag{9}$$

$$\beta_i \xi_i = 0, \quad i = 1, 2, \dots, n, \tag{10}$$

From the expression,  $\beta_i$  denotes the Lagrange Multipliers respective to  $\xi_i$  that was presented for enforcing the non-negativity of  $\xi_i$ . Where derivative of Lagrange function for primitive problem regarding  $\xi_i$  is zero, the computation of the derivative produces. Fig. 2 showcases the overview of SVM hyperplane.

$$\alpha_i + \beta_i = C. \tag{11}$$



**Figure 2:** SVM hyperplane

At the same time considered (10) and (11), we obtain;

$$\xi_i = 0 \text{ if } \alpha_i < C. \tag{12}$$

As a result, the optimum weight  $W^*$  is formulated by:

$$W^* = \sum_{i=1}^n \alpha_i^* y_i x_i. \quad (13)$$

The optimum bias  $b^*$  is attained using samples  $(x_i, y_i)$  in the trainset set where  $0 < \alpha_i^* < C$  and  $\xi_i = 0$ , as follows

$$b^* = 1 - \xi_i - \sum_{sv} \alpha_i^* y_i x_i x_s, \quad \forall y_s = +1. \quad (14)$$

### 2.3 Parameter Tuning: SSO Algorithm

To optimally modify the SM-SVM parameter values, the SSO algorithm has been exploited in this study. As the optimum hunter by their nature, the sharks are foraging nature that rotate and drives forward which is very effectual from determining prey [20]. An optimized approach for simulating shark foraging is most effectual optimized approach. To certain places, the shark transfers at speed to particles that take intensive scent, so initial velocity vector is formulated as:

$$[V_1^1, V_2^1, \dots, V_{NP}^1]$$

The sharks take inertia once it swims, so velocity equation of every dimensional is provided under,

$$V_{i,j}^k = \eta_k \cdot R_1 \cdot \left. \frac{\partial(OF)}{\partial x_j} \right|_{x_{i,j}^k} + \alpha_k \cdot R_2 \cdot v_{i,j}^{k-1} \quad (15)$$

In which  $j = (1, 2, \dots, ND)$ ,  $i = (1, 2, \dots, NP)$ , and  $k = (1, 2, \dots, k_{\max})$ ;  $ND$  implies the count of dimensional;  $NP$  signifies the count of velocity vectors (size of shark populations);  $k_{\max}$  denotes the count of iterations;  $OF$  signifies the objective function;  $\eta_k \in [0, 1]$  demonstrates the gradient co-efficient;  $\alpha_k$  implies the weighted co-efficient, therefore it can be random value amongst 0 and 1, as well  $R_1$  &  $R_2$  represented the 2 arbitrary values from range of zero's and one's. The speed of sharks is needed for preventing boundary and particular speed limitation as given below:

$$|v_{i,j}^k| = \min \left[ |v_{i,j}^k|, |\beta_k \cdot v_{i,j}^{k-1}| \right]$$

whereas  $\beta_k$  denotes speed limit factors of  $k^{th}$  iteration. The sharks take a new place  $Y_i^{k+1}$  because of moving forward, and  $Y_i^{k+1}$  is determined as the previous place and speed which is provided as:

$$Y_i^{k+1} = X_i^k + V_i^k \cdot \Delta t_k \quad (16)$$

In which  $\Delta t_k$  denotes the time interval of  $k^{th}$  iteration. Also the moving forward, sharks commonly rotate along its path for seeking strong scent particles and improve its direction of motion that is actual direction of moving.

The rotating shark moves from the closed range which could not fundamentally a circle. During the view of optimized, the shark executes local searching at all the phases to determine optimum candidate solutions. The searching equation to this place was provided under as:

$$Z_i^{k+1,m} = Y_i^{k+1} + R_3 \cdot Y_i^{k+1} \quad (17)$$

whereas  $m = (1, 2, \dots, M)$  signifies the amount of points at all the phases of place searches;  $R_3$  denotes the random value from the range of  $-1$  and  $1$ . If the shark determines a strong odor point from the rotation, it moved towards the point and endures the search direction. The position search procedure

was provided from the subsequent formula,

$$X_i^{k+1} = \arg \max \left\{ OF(Y_i^{k+1}), OF(Z_i^{k+1,1}), \dots, OF(Z_i^{k+1,M}) \right\} \quad (18)$$

As aforementioned,  $Y_i^{k+1}$  has been obtained in the linear motion and  $Z_i^{k+1,M}$  has reached in the rotational motion. The shark is choosing the candidate solution with higher computation index value as shark following place  $X_i^{k+1}$ . For boosting the convergence rate of SSO technique, the OBL model was utilized and so improves the quality of primary population solutions. It discovers either opposite or original direction solutions. The opposite number  $x$  is represented as a real value from the range of  $x \in [lb, ub]$ . The opposite number of  $x$  is determined as  $\tilde{x}$ :

$$\tilde{x} = lb + ub - x \quad (19)$$

The aforementioned formula endures normalized if it obtains executed to search region with many dimensional. For normalizing it, the search agent and the respective opposite solution is demonstrated as:

$$x = [x_1, x_2, x_3, \dots, x_D] \quad (20)$$

$$\tilde{x} = [\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \dots, \tilde{x}_D] \quad (21)$$

The value of every component from  $\tilde{x}$  is represented as:

$$\tilde{x}_j = lb_j + ub_j - x_j \text{ where } j = 1, 2, 3, \dots, D \quad (22)$$

At this point, the fitness function can be  $f(\cdot)$ . When the fitness value  $f(\tilde{x})$  of opposite solutions surpass  $f(x)$  the original solution  $x$ , then  $x = \tilde{x}$ ; else  $x = x$ .

### 3 Simulation Results and Discussion

The proposed model is simulated using Python 3.6.5 tool. The proposed model is experimented on PC i5-8600k, GeForce 1050Ti 4GB, 16GB RAM, 250GB SSD, and 1TB HDD. In this section, the sign language recognition performance of the SSODL-ASLR model is tested using a dataset comprising 3600 samples with 36 classes as illustrated in [Tab. 1](#). A few sample American Sign Language Code images are displayed in [Fig. 3](#).

**Table 1:** Dataset details

Label	Description	No. of samples
1	A	100
2	B	100
3	C	100
4	D	100
5	E	100
6	F	100
7	G	100
8	H	100
9	I	100
10	J	100

(Continued)

<b>Table 1 (continued)</b>		
Label	Description	No. of samples
11	K	100
12	L	100
13	M	100
14	N	100
15	O	100
16	P	100
17	Q	100
18	R	100
19	S	100
20	T	100
21	U	100
22	V	100
23	W	100
24	X	100
25	Y	100
26	Z	100
27	0	100
28	1	100
29	2	100
30	3	100
31	4	100
32	5	100
33	6	100
34	7	100
35	8	100
36	9	100
<b>Total number of samples</b>		<b>3600</b>

Fig. 4 implies the confusion matrix formed by the SSODL-ASLR model on the applied 70% of training (TR) data. The figure indicated that the SSODL-ASLR model has proficiently recognized all the 36 class labels on 70% of TR data.

The classifier output of the SSODL-ASLR model is derived on 70% of TR data in Tab. 2. The experimental outcomes demonstrated the SSODL-ASLR method has resulted in effectual outcomes over other models. For instance, on class label 1, the SSODL-ASLR model has provided  $ACC_Y$  of 99.29%. Moreover, on class label 10, the SSODL-ASLR model has offered  $ACC_Y$  of 99.09%. Eventually, on class label 20, the SSODL-ASLR algorithm has rendered  $ACC_Y$  of 99.01%. Meanwhile, on class label 36, the SSODL-ASLR algorithm has presented  $ACC_Y$  of 99.13%.



**Table 2:** Result analysis of SSODL-ASLR approach with distinct class labels under 70% of TR data

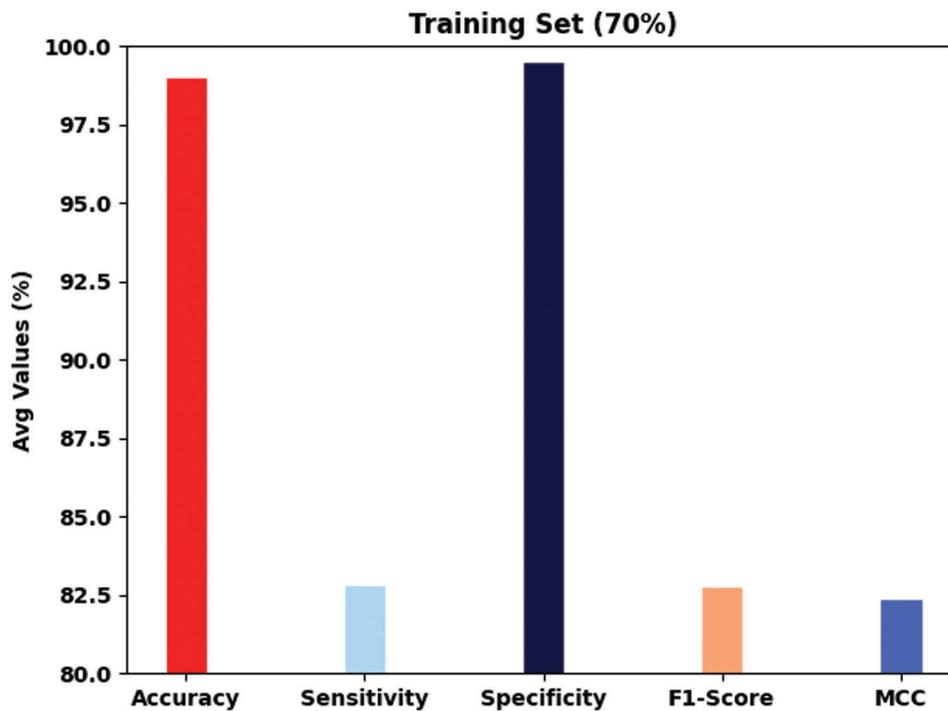
Training set (70%)					
Labels	Accuracy	Sensitivity	Specificity	F1-score	MCC
1	99.29	88.89	99.55	86.15	85.83
2	98.81	84.62	99.26	81.48	80.93
3	99.25	89.04	99.55	87.25	86.88
4	99.17	81.97	99.59	82.64	82.22
5	99.21	88.16	99.55	87.01	86.61
6	99.25	84.38	99.63	85.04	84.66
7	99.25	88.73	99.55	86.90	86.53
8	98.45	67.12	99.39	71.53	70.90
9	98.81	82.61	99.27	79.17	78.63
10	99.09	81.94	99.59	83.69	83.24
11	98.97	83.56	99.43	82.43	81.91
12	99.29	88.16	99.63	88.16	87.79
13	98.85	80.56	99.39	80.00	79.41
14	98.81	80.82	99.35	79.73	79.12
15	99.40	79.31	99.88	85.98	86.00
16	99.05	80.77	99.63	84.00	83.58
17	99.25	86.76	99.59	86.13	85.75
18	99.01	76.47	99.63	80.62	80.24
19	99.17	86.11	99.55	85.52	85.09
20	99.01	84.51	99.43	82.76	82.27
21	99.29	88.41	99.59	87.14	86.78
22	98.85	74.63	99.51	77.52	76.99
23	99.21	88.73	99.51	86.30	85.93
24	99.21	80.30	99.71	84.13	83.82
25	98.77	79.41	99.31	77.70	77.08
26	98.69	78.12	99.23	75.19	74.57
27	98.93	88.46	99.26	83.64	83.22
28	98.85	80.00	99.39	79.43	78.84
29	98.89	69.74	99.80	79.10	79.30
30	99.29	86.89	99.59	85.48	85.13
31	98.93	82.09	99.39	80.29	79.76
32	98.89	78.26	99.47	79.41	78.85
33	99.01	84.29	99.43	82.52	82.03
34	98.93	82.35	99.39	80.58	80.04

(Continued)

**Table 2 (continued)**

Training set (70%)					
Labels	Accuracy	Sensitivity	Specificity	F1-score	MCC
35	99.52	92.11	99.75	92.11	91.86
36	99.13	83.10	99.59	84.29	83.85
Average	99.05	82.82	99.51	82.81	82.38

Fig. 5 exhibits average classifier outcomes of the SSODL-ASLR model on 70% of TR data. The figure reported that the SSODL-ASLR model has reached effectual classification outcomes with average  $accu_y$  of 99.05%,  $sens_y$  of 82.82%,  $spec_y$  of 99.51%,  $F1_{score}$  of 82.81%, and Mathew Correlation Coefficient (MCC) of 83.28%.



**Figure 5:** Average analysis of SSODL-ASLR approach under 70% of TR data

Fig. 6 portrays the confusion matrix formed by the SSODL-ASLR method on the applied 30% of testing (TS) data. The figure signifies the SSODL-ASLR technique has proficiently recognized all the 36 class labels on 30% of TS data.

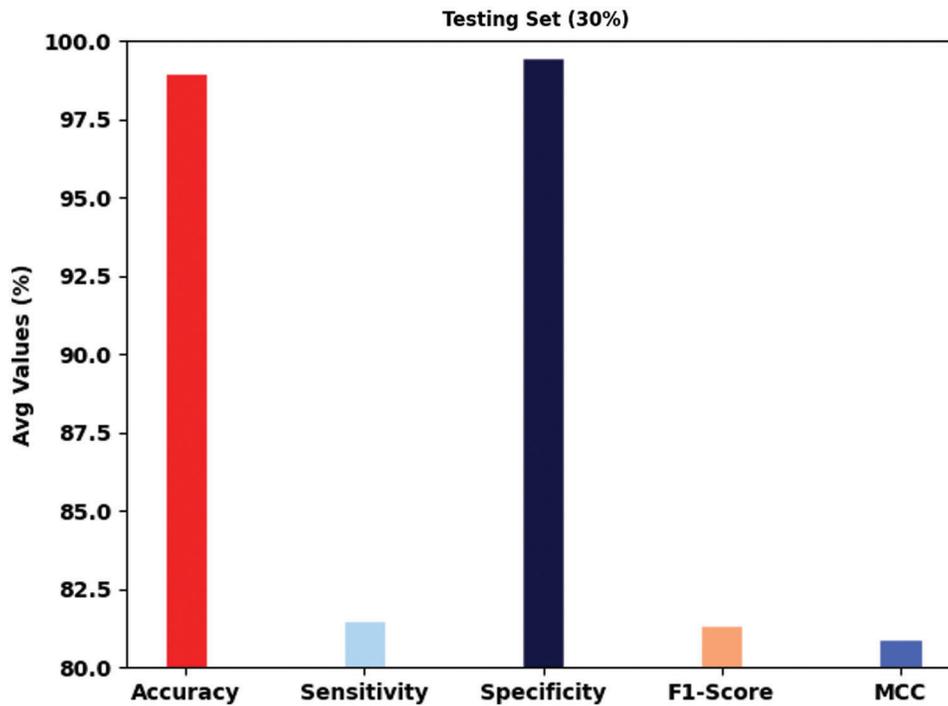
The classifier output of the SSODL-ASLR approach is derived on 30% of TS data in Tab. 3. The experimental outcomes established the SSODL-ASLR technique has resulted in effectual outcomes over other models. For example, on class label 1, the SSODL-ASLR approach has offered  $ACC_Y$  of 98.24%. Further, on class label 10, the SSODL-ASLR approach has presented  $ACC_Y$  of 98.70%. Eventually, on class label 20, the SSODL-ASLR model has granted  $ACC_Y$  of 99.17%. At the same time, on class label 36, the SSODL-ASLR model has offered  $ACC_Y$  of 99.35%.



**Table 3 (continued)**

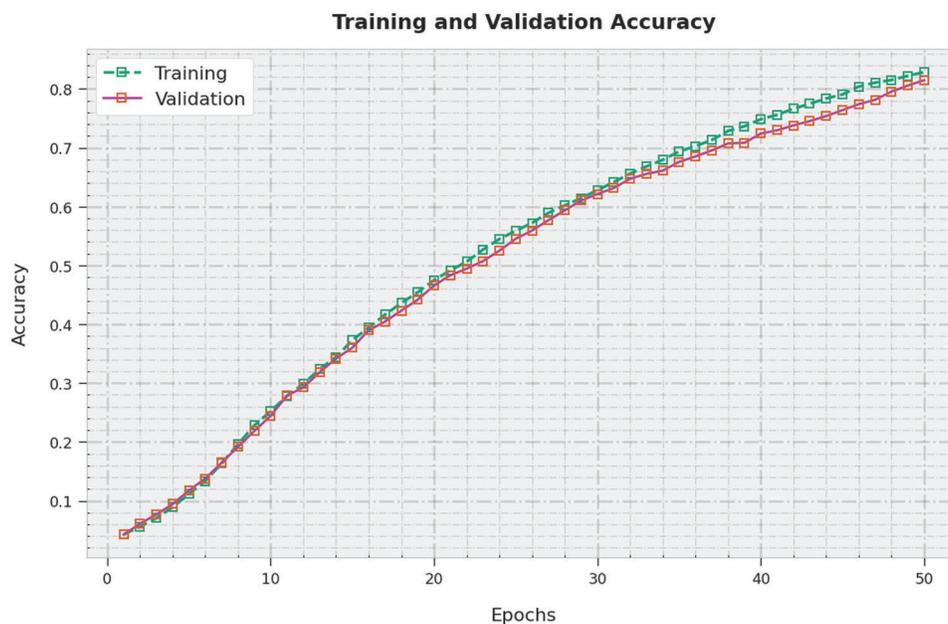
Testing set (30%)					
Labels	Accuracy	Sensitivity	Specificity	F1-score	MCC
9	98.89	83.87	99.33	81.25	80.72
10	98.70	78.57	99.24	75.86	75.24
11	99.07	81.48	99.53	81.48	81.01
12	98.98	75.00	99.53	76.60	76.09
13	98.61	78.57	99.14	74.58	73.97
14	98.61	74.07	99.24	72.73	72.03
15	98.70	78.57	99.52	82.50	81.94
16	99.72	95.45	99.81	93.33	93.22
17	99.35	90.62	99.62	89.23	88.91
18	99.44	93.75	99.62	90.91	90.67
19	99.07	78.57	99.62	81.48	81.07
20	99.17	79.31	99.71	83.64	83.34
21	98.80	77.42	99.43	78.69	78.08
22	98.98	87.88	99.33	84.06	83.62
23	99.44	82.76	99.90	88.89	88.86
24	98.61	79.41	99.24	78.26	77.55
25	99.07	81.25	99.62	83.87	83.44
26	99.07	86.11	99.52	86.11	85.63
27	99.07	81.82	99.43	78.26	77.87
28	99.26	86.67	99.62	86.67	86.29
29	98.89	70.83	99.53	73.91	73.42
30	98.70	79.49	99.42	81.58	80.94
31	99.07	81.82	99.62	84.38	83.94
32	98.98	80.65	99.52	81.97	81.45
33	99.07	90.00	99.33	84.38	84.07
34	98.70	84.38	99.14	79.41	78.89
35	99.07	79.17	99.53	79.17	78.69
36	99.35	86.21	99.71	87.72	87.40
Average	98.97	81.48	99.47	81.37	80.91

Fig. 7 displays an average classifier outcome of the SSODL-ASLR model on 30% of TS data. The figure reported that the SSODL-ASLR model has reached effectual classification outcomes with average  $accu_y$  of 98.97%,  $sens_y$  of 81.48%,  $spec_y$  of 99.47%,  $F1_{score}$  of 81.37%, and MCC of 80.91%.



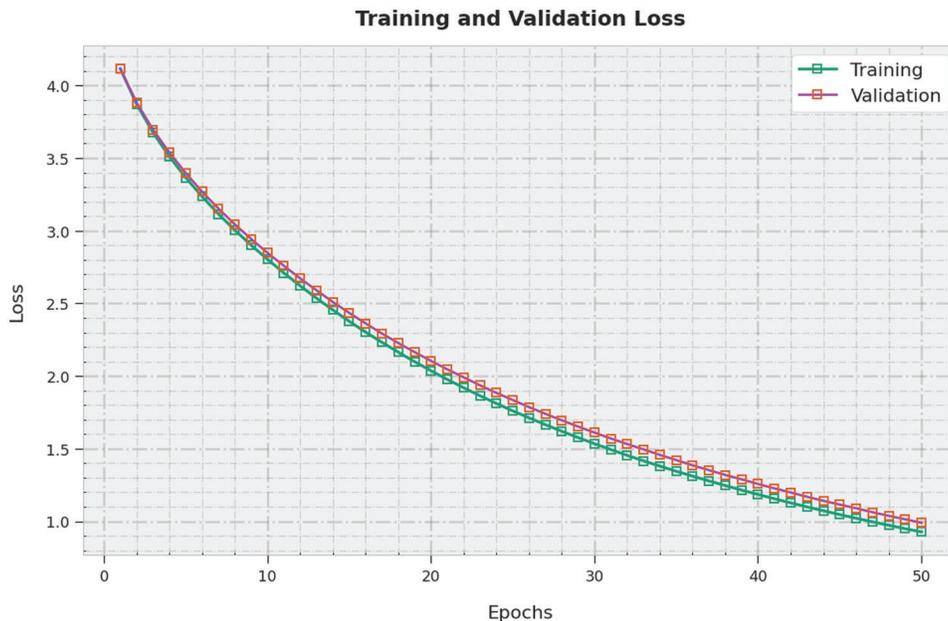
**Figure 7:** Average analysis of SSODL-ASLR approach under 30% of TS data

The training accuracy (TA) and validation accuracy (VA) acquired by the SSODL-ASLR method on test dataset is established in Fig. 8. The experimental outcome inferred the SSODL-ASLR technique has achieved maximal values of TA and VA. Predominantly the VA is greater than TA.



**Figure 8:** TA and VA analysis of SSODL-ASLR approach

The training loss (TL) and validation loss (VL) attained by the SSODL-ASLR approach on test dataset are established in Fig. 9. The experimental outcome implied the SSODL-ASLR algorithm has exhibited least values of TL and VL. In specific, the VL is lesser than TL.



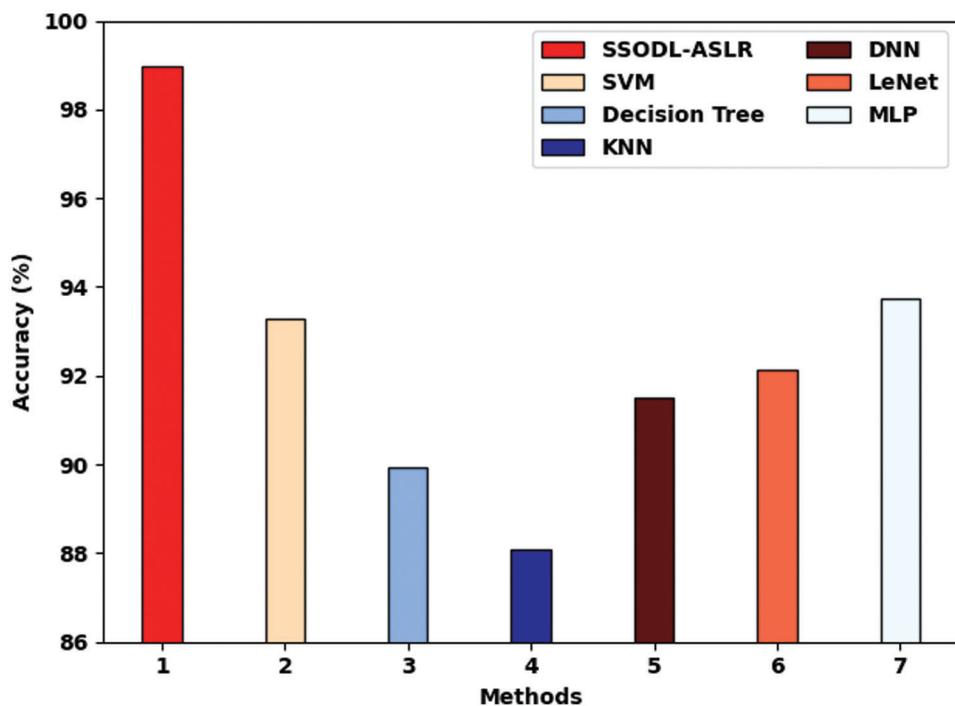
**Figure 9:** TL and VL analysis of SSODL-ASLR approach

Tab. 4 exemplifies a comparative assessment of the SSODL-ASLR model with recent models [21]. Fig. 10 reports a brief comparison study of the SSODL-ASLR model with latest methods interns of  $ACC_{U_Y}$ . The results pointed out that the SSODL-ASLR model has offered higher  $ACC_{U_Y}$  of 98.97% whereas the SVM, decision tree (DT), k-nearest neighbour (KNN), DNN, LeNet, and multilayer perceptron (MLP) models have exhibited ineffectual performance with lower  $ACC_{U_Y}$  values.

**Table 4:** Comparative analysis of SSODL-ASLR approach with recent methodologies

Methods	Accuracy	Sensitivity	Specificity
SSODL-ASLR	98.97	81.48	99.47
SVM	93.28	79.45	96.07
Decision tree	89.92	76.59	93.15
KNN	88.07	75.03	96.30
DNN	91.52	77.72	93.23
LeNet	92.14	74.64	92.90
MLP	93.72	77.63	97.90

The results highlight the SSODL-ASLR model has granted higher  $Sens_y$  and  $Spec_y$  of 81.48% and 99.47% whereas the SVM, DT, KNN, DNN, LeNet, and MLP models have displayed ineffectual performance with lower  $Sens_y$  and  $Spec_y$  values. These results and discussion reported that the SSODL-ASLR model has shown effectual performance over other ML and DL models.



**Figure 10:** *Accu<sub>y</sub>* analysis of SSODL-ASLR approach with existing methodologies

#### 4 Conclusion

In this study, a new SSODL-ASLR technique was introduced for the recognition of sign language for hearing and speaking impaired people. The presented SSODL-ASLR technique majorly concentrates on the recognition and classification of sign language provided by deaf and dumb people. The presented SSODL-ASLR model encompasses a two stage process namely sign language detection and sign language classification. Primarily, the Mask RCNN model is exploited for sign language recognition. In the next stage, the SSO algorithm with SM-SVM model is utilized for sign language classification. To assure the enhanced classification performance of the SSODL-ASLR model, a brief set of simulations was carried out. The extensive results portrayed the supremacy of the SSODL-ASLR model over other techniques.

**Funding Statement:** The author received no specific funding for this study.

**Conflicts of Interest:** The author declares that he has no conflicts of interest to report regarding the present study.

#### References

- [1] M. A. Ahmed, B. B. Zaidan, A. A. Zaidan, M. M. Salih and M. M. bin Lakulu, "A review on systems-based sensory gloves for sign language recognition state of the art between 2007 and 2017," *Sensors*, vol. 18, no. 7, pp. 2208, 2018.
- [2] M. Friedner, "Sign language as virus: Stigma and relationality in urban India," *Medical Anthropology*, vol. 37, no. 5, pp. 359–372, 2018.
- [3] X. Jiang, S. C. Satapathy, L. Yang, S. -H. Wang and Y. -D. Zhang, "A survey on artificial intelligence in Chinese sign language recognition," *Arabian Journal for Science and Engineering*, vol. 45, no. 12, pp. 9859–9894, 2020.

- [4] V. S. Pineda and J. Corburn, "Correction to: Disability, urban health equity, and the coronavirus pandemic: Promoting cities for all," *Journal of Urban Health*, vol. 98, no. 2, pp. 308–308, 2021.
- [5] M. Al Duhayyim, H. Mesfer Alshahrani, F. N. Al-Wesabi, M. Abdullah Al-Hagery, A. Mustafa Hilal *et al.*, "Intelligent machine learning based EEG signal classification model," *Computers, Materials & Continua*, vol. 71, no. 1, pp. 1821–1835, 2022.
- [6] P. Kumar, R. Saini, P. P. Roy and D. P. Dogra, "A position and rotation invariant framework for sign language recognition (SLR) using Kinect," *Multimedia Tools and Applications*, vol. 77, no. 7, pp. 8823–8846, 2018.
- [7] S. A. Qureshi, S. E. Raza, L. Hussain, A. A. Malibari, M. K. Nour *et al.*, "Intelligent ultra-light deep learning model for multi-class brain tumor detection," *Applied Sciences*, 2022. <https://doi.org/10.3390/app12083715>.
- [8] N. S. Khan, A. Abid and K. Abid, "A novel natural language processing (nlp)-based machine translation model for English to Pakistan sign language translation," *Cognitive Computation*, vol. 12, no. 4, pp. 748–765, 2020.
- [9] M. Khari, A. K. Garg, R. G. Crespo and E. Verdú, "Gesture recognition of rgb and rgb-d static images using convolutional neural networks," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 7, pp. 22, 2019.
- [10] S. Shivashankara and S. Srinath, "American sign language recognition system: An optimal approach," *International Journal of Image, Graphics and Signal Processing*, vol. 11, no. 8, pp. 18, 2018.
- [11] J. G. Ruiz, C. M. T. González, A. T. Fettmilch, A. P. Roescher, L. E. Hernández *et al.*, "Perspective and evolution of gesture recognition for sign language: A review," *Sensors*, vol. 20, no. 12, pp. 3571, 2020.
- [12] N. B. Ibrahim, H. H. Zayed and M. M. Selim, "Advances, challenges and opportunities in continuous sign language recognition," *Journal of Engineering and Applied Science*, vol. 15, no. 5, pp. 1205–1227, 2019.
- [13] A. Sharma, N. Sharma, Y. Saxena, A. Singh and D. Sadhya, "Benchmarking deep neural network approaches for Indian sign language recognition," *Neural Computing and Applications*, vol. 33, no. 12, pp. 6685–6696, 2021.
- [14] J. J. Bird, A. Ekárt and D. R. Faria, "British sign language recognition via late fusion of computer vision and leap motion with transfer learning to American sign language," *Sensors*, vol. 20, no. 18, pp. 5151, 2020.
- [15] D. A. Kumar, A. S. C. S. Sastry, P. V. V. Kishore and E. K. Kumar, "Indian sign language recognition using graph matching on 3D motion captured signs," *Multimedia Tools and Applications*, vol. 77, no. 24, pp. 32063–32091, 2018.
- [16] L. N. Zeledón, J. Peral, A. Ferrández and M. C. Rivas, "A systematic mapping of translation-enabling technologies for sign languages," *Electronics*, vol. 8, no. 9, pp. 1047, 2019.
- [17] K. Parvez, M. Khan, J. Iqbal, M. Tahir, A. Alghamdi *et al.*, "Measuring effectiveness of mobile application in learning basic mathematical concepts using sign language," *Sustainability*, vol. 11, no. 11, pp. 3064, 2019.
- [18] I. P. Borrero, D. M. Santos, M. J. V. Vazquez and M. E. G. Arias, "A new deep-learning strawberry instance segmentation methodology based on a fully convolutional neural network," *Neural Computing & Applications*, vol. 33, no. 22, pp. 15059–15071, 2021.
- [19] H. Liang, H. Bai, N. Liu and X. Sui, "Polarized skylight compass based on a soft-margin support vector machine working in cloudy conditions," *Applied Optics*, vol. 59, no. 5, pp. 1271, 2020.
- [20] L. Wang, X. Wang, Z. Sheng and S. Lu, "Multi-objective shark smell optimization algorithm using incorporated composite angle cosine for automatic train operation," *Energies*, vol. 13, no. 3, pp. 714, 2020.
- [21] T. Chong and B. Lee, "American sign language recognition using leap motion controller with machine learning approach," *Sensors*, vol. 18, no. 10, pp. 3554, 2018.