Mathematics

AIMS Mathematics, 10(11): 25154–25174.
DOI: 10.3934/math.20251113
Received: 06 August 2025
Revised: 22 September 2025
Accepted: 17 October 2025
Published: 03 November 2025

https://www.aimspress.com/journal/Math

Research article

Innovative sign language accessibility technique for hearing and speech impaired: deep learning-based hand gesture recognition for communication

Najm Alotaibi[1,2,*], Alanoud Subahi[3], Nouf Atiahallah Alghanmi[3] and Mohammed Rizwanullah[4]

1 Prince Saud AlFaisal Institute for Diplomatic Studies, Riyadh, Saudi Arabia
2 King Salman Centre for Disability Research, Riyadh 11614, Saudi Arabia
3 Department of Information Technology, Faculty of Computing and Information Technology, King Abdulaziz University, Rabigh 25732, Saudi Arabia
4 Department of Computer and Self Development, Preparatory Year Deanship, Prince Sattam bin Abdulaziz University, AlKharj, Saudi Arabia

* **Correspondence:** Email: alotaibinajm@gmail.com.

Abstract: Sign language (SL) plays a significant part in communication among people who are hearing and deaf. Silent people struggle to convey their message to others. Since most people have not received a formal language education, it is highly complex to transfer messages in an emergency. Hence, a solution to this problem is to convert SL into a human voice. Gesture-to-speech systems usually use either vision-based or non-vision-based technologies, such as cameras or wearable sensors. However, many existing solutions lack cost-effectiveness and flexibility; for example, some depend on specific hardware or only function in controlled environments. In this paper, the Advancing Sign Language Accessibility using Deep Learning-Based Hand Gesture Recognition (ASLA-DLHGR) technique for hearing and speech-impaired individuals is proposed. The goal of the ASLA-DLHGR technique is to recognize hand gestures for communication among disabled people. Initially, the data pre-processing process is performed using the bilateral filtering (BF) model. Furthermore, the ASLA-DLHGR technique employs the SqueezeNet model to learn composite features from the pre-processed data. Moreover, the tunicate swarm algorithm (TSA) based hyperparameter process is performed to enhance the performance of the SqueezeNet method. For the gesture recognition process, a hybrid of a convolutional neural network and a bidirectional long short-term memory (CNN-BiLSTM) method is implemented. To demonstrate the managed gesture recognition proficiency of the ASLA-DLHGR method, a comprehensive comparative study is carried out under the American SL dataset. The comparison study of the ASLA-DLHGR method portrayed a superior accuracy value of 99.98% over existing models.

Keywords: hand gesture recognition; sign language; hearing and speech impaired; tunicate swarm algorithm; SqueezeNet
Mathematics Subject Classification: 37M10

1. Introduction

Communication is an essential aspect of human life. People convey their thoughts and desires through speech, while others listen and respond accordingly [1]. However, some people are unable to talk and cannot even hear. It is a significant challenge for these societies to communicate. Sign language (SL) is used as a communication method between and among deaf and hearing people [2]. Since most people may not understand SL, they seek a qualified individual to help them understand it. Hand gestures are used as a way for people to express feelings and thoughts, thus helping to reinforce the information conveyed in daily conversations [3]. SL is a structured system of hand gestures, which integrates visual signs and motions, and is utilized as a means of communication. Instead of the speech-impaired and deaf community, SL is a valuable tool for daily interactions. SL encompasses procedures for various body parts, including the arms, hands, body, facial expressions, fingers, and head, to convey information [4]. However, SL is rare among the hearing community, and someone who can understand it is even rarer. This poses an honest communication obstacle between the rest of society and the deaf community, a problem that remains unsolved to this day. SL and gesture detection encompasses the entire process of recognizing and tracking signals, altering and executing them into meaningful expressions and words [5].

SL recognition refers to the use of models and procedures that identify the resultant sequence of gestures and clarify their meaning in a language or manuscript [6]. This method encompasses various research fields, including pattern natural language processing (NLP), video acquisition, detection, computer vision (CV), human-computer interaction (HCI), and more [7]. It is a demanding topic with the highest complexity. At present, the usual SL recognition methods are mainly divided into two types: machine vision and sensor-based systems. The existing technologies are not adequately described, as they fail to address specific issues such as lighting issues, wearable sensor inconvenience, and poor real-time performance, which makes the research justification unclear [8]. Artificial intelligence (AI) methods are utilized for processes that enable computers to display human-like intelligent actions, such as decision-making, visual perception, natural language understanding, and speech detection [3]. Particular applications of AI include expert systems, machine vision, speech detection, and NLP. A method that can display SL is exposed by the adaptive interface between blind and deaf individuals using voice processing and hand gesture recognition, which allows people with normal hearing to interconnect with talk-impaired individuals or groups of people even better [9].

In this paper, the Advancing Sign Language Accessibility using Deep Learning-Based Hand Gesture Recognition (ASLA-DLHGR) technique for hearing and speech-impaired individuals is proposed. The goal of the ASLA-DLHGR technique is to recognize hand gestures for communication among disabled people. Initially, the data pre-processing process is performed using the bilateral filtering (BF) model. Furthermore, the ASLA-DLHGR technique employs the SqueezeNet model to learn composite features from the pre-processed data. Moreover, the tunicate swarm algorithm (TSA) based hyperparameter process is performed to enhance the performance of the SqueezeNet method. For the gesture recognition process, a hybrid of a convolutional neural network and a bidirectional long short-term memory (CNN-BiLSTM) method is implemented. To demonstrate the managed gesture recognition proficiency of the ASLA-DLHGR method, a comprehensive comparative study is performed using the American SL dataset. The key contributions of the ASLA-DLHGR method is listed below.

- The BF model is utilized in the pre-processing process to enhance the gesture data quality by mitigating noise while preserving significant edge details. This step ensures a cleaner input for the feature extraction stage. Additionally, it improves the model's capability to handle discrepancies in gesture input. As a result, it contributes to more robust and accurate gesture recognition.
- The effective and compact SqueezeNet technique is employed to extract meaningful spatial features from gesture data while reducing computational load. This facilitates faster processing without compromising feature quality. Additionally, it assists deployment on resource-constrained devices. This improves the practicality of the model for real-time gesture recognition applications.
- The hybrid deep learning (DL) methodology which utilizes the CNN technique is employed to extract spatial features, and BiLSTM is used to capture temporal dependencies in gesture sequences. This integration effectually models both spatial and sequential patterns. It also enhances the recognition accuracy for dynamic gestures and improves the ability of the model to interpret intrinsic gesture inputs in real time.
- The incorporation of BF, lightweight SqueezeNet, and a CNN-BiLSTM hybrid presents an efficient end-to-end framework for gesture recognition. The spatial and temporal challenges of the model are efficiently addressed, while also maintaining low model complexity. This model plays a significant role in assisting real-time applications with limited computational resources. The novelty is in integrating these techniques to achieve high accuracy with minimal overhead.

2. Literature review

Abdul Ameer et al. [10] introduced a long short-term memory (LSTM) model in combination with MediaPipe to mitigate the hurdles and effectively communicate and connect deaf individuals. The methodology integrates an attention mechanism and an LSTM for processing the extracted key points from captured signs. The attention layer focuses on relevant sectors of the given order, while the LSTM manages temporal associations and encodes the sequential information. The authors [11] proposed a novel methodology of FFNN for automated detection. This system detects hand signals by extracting feature points using the FFNN technique. Hand gesture recognition (HGR) with voice processing by using a Hidden Markov model (HMM) is implemented to ease individuals with hearing issues. Valarmathi et al. [12] a technique that incorporated NLP methodologies to enhance the articulacy and coherence of the decoded text. Furthermore, expressive 3D gestures are employed to animate the SL, making the communication more engaging and relevant. These gestures are customizable to match the user's identity, further personalizing the communication. This system leverages NLP, DL, and 3D to address message barriers for individuals who have difficulties with speech and hearing. Jebali et al. [13] propose a manual and non-manual (MNM) technique. This technique employs a CNN, known as VGG16net, to utilize a methodology based on training on the video dataset, implementing the

Multimodal Spatial and Temporal Representation (MSR and MTR) of diverse models. This approach outlines temporal alterations from non- and reliant pathways to examine the cooperation of several models. A cooperative optimizer technique, abstracted by the utilization of a multi-scale perception module, is also employed. Miah et al. [14] introduced a dual-stream multi-phase graph convolution with attention and residual connection (GCAR) constructed for MSR-related data. The presented approach, integrating a channel attention component, improves attention levels, specifically for non-related skeleton points at the time of particular events under MTR factors. Bhatt and Dash [15] proposed an advanced DL-based approach. This dataset is pre-processed to optimize its relevance for ensuing training of a customized CNN methodology. The trained CNN approach portrayed excellence in detecting and interpreting real-time SL signs. Specifically, real-time recognition of American SL (ASL) implemented Teachable Machine and MediaPipe, whereas the following technique demonstrated real-time hand gesture detection by employing a convexity-assisted method. Shin et al. [16] integrated joint skeleton-assisted handcrafted factors and pixel-assisted transfer learning (TL) technique, namely ResNet101. This approach is comprised of two discrete feature extraction cstreams: initially, significant handcrafted factors are extracted, underscoring the capture of hand orientation data within KSL signs. subsequently, a DL-based ResNet-101 is implemented for capturing hierarchical representations of the KSL alphabet sign. Finally, the integrated feature is sent to the DL-based classification module for the classification process.

Jayasingh, Rani, and Swathi [17] developed a lightweight CNN based SL translator to recognize gestures. Additionally, the system also utilizes advanced filtering techniques and neural network classifiers, specifically Visual Geometry Group 19 (VGG19) and Residual Network 50 (ResNet50) models with MediaPipe. Aurangzeb et al. [18] introduced a novel hand vision-based CNN model (HVCNNM) methodology. Elgohr et al. [19] proposed a real-time ASL recognition system that utilizes the You Only Look Once version 11 (YOLOv11) model with improved architecture and training techniques. Rathnayake et al. [20] presented a real-time gesture detection system by utilizing a glove with flex sensors and ML methods, including support vector machine (SVM), k-nearest neighbors (kNN), and naïve bayes (NB) techniques. Malviya, Mahajan, and Sethi [21] developed an accurate and reliable Indian Sign Language (ISL) recognition system to interpret SL gestures and assisting assistive technologies for the deaf community. Miah et al. [22] developed GmTC, an end-to-end sign language recognition (SLR) system that utilizes a graph convolutional network (GCN) and an attention-based DL method to translate multi-cultural SL into text accurately. Singhal et al. [23] presented the Dumb Aid Phone system to recognize hand gestures and convert them into speech, enabling effective communication for deaf and non-speaking individuals. Rehman et al. [24] developed a deep CNN (DCNN) version 2 technique to assist hearing-impaired individuals. Soukaina, Mohammed, and Mohamed [25] developed a lightweight ML technique by utilizing novel geometrical features from hand landmarks. Kukreja, Singh, and Chauhan [26] proposed a gesture vision (GV) technique, a real-time static SLR system that utlize MediaPipe, OpenCV, and a random forest (RF) classifier to recognize diverse hand gestures.

The limitations of the existing studies include a dependence on intrinsic DL models that demand high computational resources, making real-time deployment on low-power devices challenging. Many approaches concentrate on specific SLs or static gestures, restricting generalizability across multi-cultural or dynamic SL. Several models lack robustness against discrepancies in backgrounds, lighting, and occlusions. Additionally, there is an insufficient exploration of efficient feature fusion methods that balance accuracy and model size. Furthermore, most systems depend on massive annotated datasets, which are often scarce for less common sign languages. There is a need to develop lightweight, scalable, and adaptable models capable of real-time recognition across diverse SLs while maintaining high accuracy under varying environmental conditions and limited data availability.

3. Proposed methodology

In this article, the ASLA-DLHGR model is proposed for individuals with hearing and speech impairments. The goal of the ASLA-DLHGR technique is to recognize hand gestures to communicate among disabled individuals. To accomplish that, the ASLA-DLHGR technique is comprised four processes, pre-processing, feature extraction, parameter choice, and classification process. Figure 1 portrays the workflow of the ASLA-DLHGR technique.

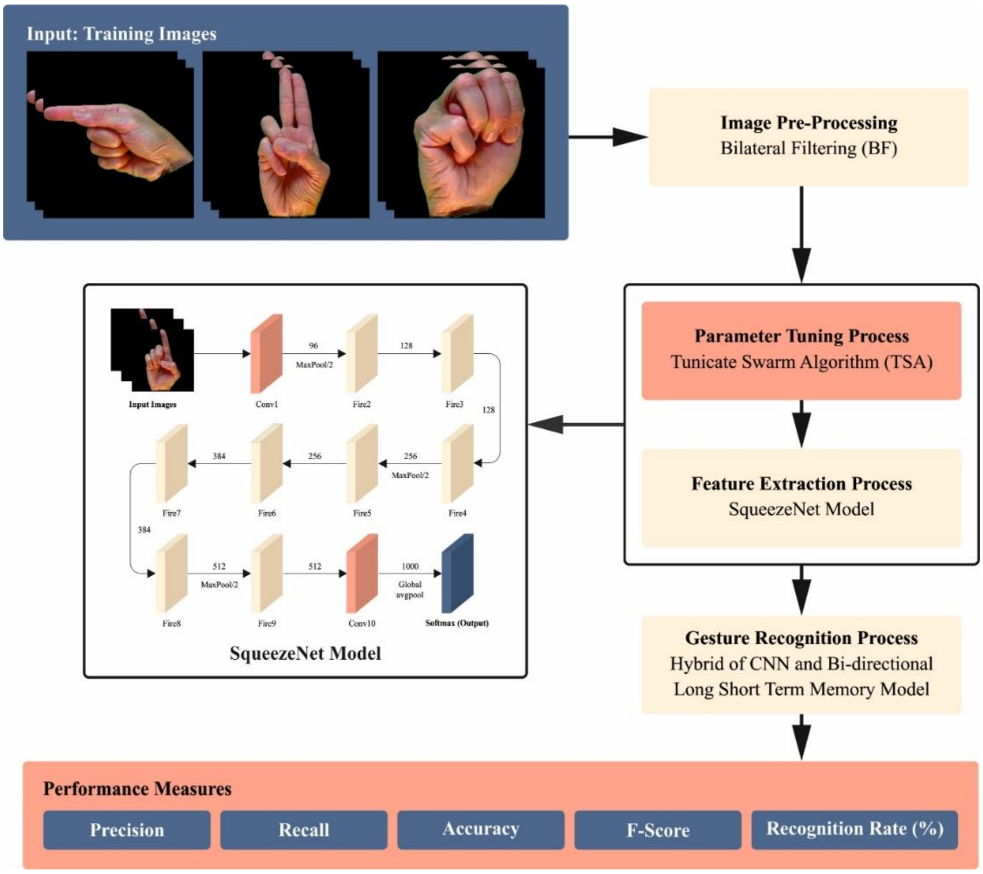

Figure 1. Workflow of ASLA-DLHGR technique.

3.1. Image pre-processing: BF model

At the primary level, the ASLA-DLHGR technique utilizes BF to perform pre-processing [27]. The model shows excellence in smoothening images while also preserving crucial edge details that are considered significant for gesture boundary clarity. The BF technique also maintains spatial structure, improving feature extraction in later stages, unlike conventional filters, namely Gaussian or median filtering (MF), that often blur edges. This is highly appropriate for the recognition process, where fine edge details are crucial. This technique also effectually handles noise, enhancing the quality of input data without distorting gesture shapes. Its balance between noise reduction and edge preservation makes it more appropriate than other filtering methods. Overall, BF strengthens the robustness and accuracy of the entire recognition process.

BF is an excellent image pre-processing method that improves the qualities of hand gesture images utilized in SL detection for hearing and speech-impaired individuals. With smooth images, with protective edges, BF successfully decreases noise and increases the visibility of vital features like hand movements and shapes. This pre-processing stage is critical to enhance the accuracy of the following gesture recognition methods. It permits richer differences between intelligent hand gestures, enabling improved communication for users. Generally, BF plays an essential part in enhancing the efficiency of SL recognition methodologies.

3.2. Feature extractor: SqueezeNet method

ASLA-DLHGR technique applies the SqueezeNet model to learn composite features from the pre-processed data [28]. This model effectively learns composite features from the pre-processed data due to its capability in achieving high accuracy with significantly fewer parameters. The model is considered highly appropriate due to its lightweight architecture, particularly on devices with restricted computational resources. This methodology utilizes Fire modules to mitigate the model size without losing performance, thus facilitating faster inference and lower memory usage. This technique also allows efficient training and deployment in edge computing environments. The scalability of the model is ensured by its compact design, while also maintaining effective spatial feature extraction. Thus, this is considered more appropriate for resource-efficient DL model applications.

The SqueezeNet CNN structure was introduced to reach higher accuracy with a very compact model size. Numerous developments allow SqueezeNet to reduce parameter counts. The model utilizes two Fire modules, the expand and squeeze convolution layers. The squeeze layer uses $1x1$ filters, while the expand layer combines an amalgamation of $1x1$ and $3x3$ convolution filters to implement channel-to-channel squeezing, which ensures nothing but a decrease in the filter counts in the squeeze layers. Although down-sampling operations may decrease accuracy, they can improve it based on their position in the network. In SqueezeNet, downsampling is purposefully suspended until the next phase, enabling convolutional layers to process larger activation maps. This approach allows for better accuracy with smaller filters and further facilitates efficient model compression methods, such as quantizing weights to 8-bit precision, which ultimately results in a significantly smaller model's overall disk footprint.

3.3. Hyperparameter selection: TSA technique

In addition, the TSA-based hyperparameter process is performed to improve the performance of the SqueezeNet method [29]. This model presents an improved convergence speed and solution quality in complex, multi-dimensional search spaces and is also robust in its global search capability and ability to escape local optima. It also dynamically balances exploration and exploitation, which results in more optimal hyperparameter settings, unlike conventional grid or random search models. This results in an enhanced model accuracy, stability, and generalization. The adaptive behaviour and efficiency of the TSA technique make it an ideal choice over conventional optimization techniques, specifically for DL methods such as SqueezeNet that require fine-tuned parameters for best performance.

TSA is a new bioinspired MHA that simulates the social searching behavior of bioluminescent tunicates that exist in the deep ocean. Every tunicate is cylinder-shaped and shows a gelatinous tunic that helps the tunicates to communicate with each other. These tunicates use social intelligence and water planes to determine the location of food. Over the atrium's siphons, every tunicate might quickly eject the formerly inhaled seawater, which generates a kind of jet propulsion that moves forward abruptly. Still, tunicates show swarm intelligence by transferring search information about the food source. To express the computation formulation of the jet propulsion method, it's essential for a tunicate to fulfil the following restrictions.

On the other hand, the tunicate's swarming behavior permits the searched individual to communicate location information amongst one another. This mechanism helps in upgrading the location of the tunicates depending on the optimal solutions. The mathematical expressions of these three methods are presented in the following subcategories:

3.3.1 Preventing collisions among the search individuals

To stop collisions among the search individuals, other tunicates, the following mechanisms are used to compute the updated location of the searched individual:

$$\vec{A} = \frac{\vec{G}}{\vec{M}}, \tag{1}$$

$$\vec{G} = r_2 + r_3 - \vec{F}, \tag{2}$$

$$\vec{F} = 2 \cdot r_1, \tag{3}$$

whereas the vector $\vec{A}$ was applied to define the novel location of every tunicate, the vectors $\vec{F}$ and $\vec{G}$ specify the water flow rate and the gravitational force in the deep ocean, respectively, and r_1, r_2, and r_3 are arbitrary numbers distributed ranging between (0-1). The vector $\vec{M}$ specifies the social powers among the search individuals.

$$\vec{M} = \lfloor P_{\min} + r_1 \cdot (P_{\max} - P_{\min}) \rfloor, \tag{4}$$

where, $P_{\min}$ and $P_{\max}$ are fixed to one and four, respectively, and designate the initial and secondary speediness of the search individuals to enable social interaction.

3.3.2 Moving toward the direction of the best search individual

After completing the previous step, all must continue near the path of the better search individual. The mathematical presentation to approach the best search individual is described as:

$$\vec{SD} = |F_{best} - rand \cdot x_i(t)|, \tag{5}$$

where, the vector $\vec{SD}$ denotes the spatial distance from the tunicate to the food source, F_{best}

represents the optimal food location, $x_i(t)$ symbolizes the location of the i^{th} tunicate at iteration t, and $rand \in [0,1]$.

3.3.3 Converge to the region neighbouring the better search individual

To guarantee the search individual's behaviour sufficient local exploration in closer proximities to the location of the best search individual, their locations are assessed by Eq (6).

$$x_i(t) = \begin{cases} F_{best} + \vec{A} \cdot \vec{SD}, if\ rand \geq 0.5 \\ F_{best} - \vec{A} \cdot \vec{SD}, if\ rand < 0.5 \end{cases}. \tag{6}$$

Every tunicate searches the area enclosed by the F_{best} at the tth iteration and allocates the discoveries to $x(t)$ for updating its position.

3.3.4 Tunicate's swarming behaviour

During this swarm intellect mechanism, the locations of tunicates are upgraded depending on the locations of the primary dual finest tunicates. These behaviours are demonstrated as shown:

$$x_i(t+1) = \begin{cases} \frac{x_?(t)+x_{?-1}(t+1)}{2+r_1}, & if\ i > 1 \\ x_i(t), & if\ i = 1 \end{cases}. \tag{7}$$

Here $i = 1,2, \dots n,$ n denotes size of population of the tunicates, $x_i(t+1)$ refers to upgraded location of recent search individual in the following iteration, $x_{i-1}(t+1)$ represents location of previous search individual of the following iteration, and $x_i(t)$ is defined by Eq (6). Furthermore, the visual representation of upgrading the location of every tunicate in relation to the location $x_i(t)$ is described.

The TSA develops a fitness function (FF) to identify an optimal classifier solution. It decides a positive value to emulate the best efficiency of candidate outcomes. In such research, the reduction in the classification error ratio is perceived as FF.

$$fitness(x_i) = ClassifierErrorRate(x_i)$$

$$= \frac{No.of\ misclassified\ instances}{Total\ no.of\ instances} \times 100. \tag{8}$$

3.4. Classification process: hybrid CNN-BiLSTM

For the gesture recognition process, the hybrid CNN-BiLSTM method is applied [30]. This method is chosen for its capability to effectively capture both spatial and temporal features in gesture sequences. The CNN model excels at extracting local spatial features from image frames, whereas BiLSTM effectively handles long-range dependencies and temporal dynamics across the sequence. The CNN and RNN models exhibit limitations by losing temporal context and difficulty in long-term dependencies, which this hybrid model addresses. The BiLSTM technique processes data in both forward and backward directions, enhancing context understanding. Thus, higher recognition accuracy and robustness are ensured in dynamic gesture interpretation compared to single-network approaches.

Figure 2 represents the architecture of CNN-BiLSTM techniques.

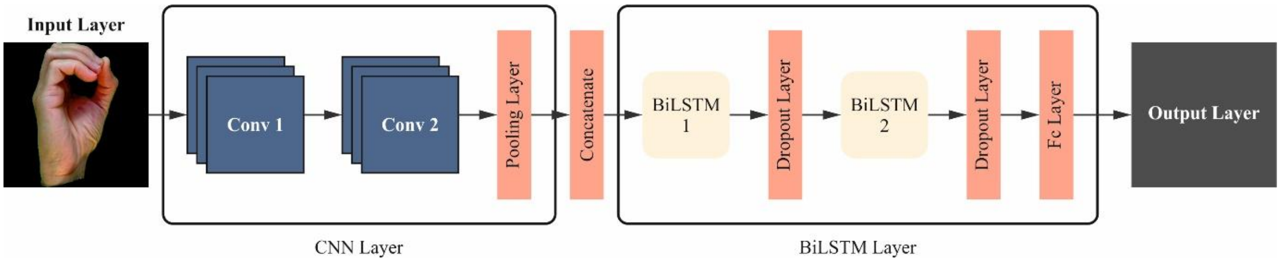

Figure 2. Architecture of CNN-BiLSTM.

A CNN is a generally applied DL algorithm in machine vision and is commonly employed in AI. CNNs are comprised of 1-, 2-, and 3-dimensional CNNs that are used to handle sequential images, videos, and signals, respectively. In the architecture, the input network existed as an image sequence after feature removal, and the programming data needed to be similar to the present framework, which improved the association between the complete structure and real productions. The CNN strategy was stimulated by the human visual method, particularly the visual cortex performance. This method executes mechanisms like weight sharing, localized perceptual domains, and so on. The fundamental architecture of a CNN comprises activation, pooling, convolutional, and fully connected (FC) layers.

CNNs can efficiently remove local features from the sequence of input images and associate them with data, but they suffer from the difficulty of losing information after handling sequential data. To resolve this difficulty, RNNN (LSTM, GRU, and RNN) is presented specially for removing time series features from sequential data. RNNs obtain input and make output at every time step by introducing a recurrent architecture; however, they maintain a hidden layer (HL) in which data from preceding time steps is saved and passed. It enables RNNs to model sequential and contextual relations in sequential data efficiently. Nevertheless, if the sequence of input is longer, the RNN updating is understood by following matrix multiplications. Hence, the operation of matrix multiplication in BP might result in the difficulty of gradient explosion or gradient vanishing. The LSTM network successfully resolves the challenges present in RNNs by presenting a gating process. The LSTM network absorbs longer and short-term time series characteristics of sequential data by guiding the weighting of input, output, and forgetting gates. It is appropriate for the classification and prediction of longer sequence data. The memory cell parameters are upgraded at every tth moment.

$$i_t = \sigma(W_{xi}\mathrm{x}_\mathrm{t} + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i), \tag{9}$$

$$f_t = \sigma\left(W_xfx_t + W_hfh_{t-1} + W_cfc_{t-1} + b_f\right), \tag{10}$$

$$c_t = f_tc_{t-1} + i_t\tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c), \tag{11}$$

$$\mathrm{o}_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_\mathrm{o}), \tag{12}$$

$$h_t = o_t\tanh(c_t), \tag{13}$$

whereas b_f and w_f represent biases and weights of the forgetting gate, x_i denotes present time input, σ refers to a function of sigmoid, and h_t signifies HL.

The Bi-LSTM network is an expansion of the LSTM network that presents a bi-directional architecture based on LSTM.

$$h_t = \sigma(w_1 x_t + w_2 h_{t-1}) \times \tanh(C_t), \tag{14}$$

$$h'_t = \sigma(w_3 x_t + w_4 h'_{t-1}) \times \tanh(C'_t), \tag{15}$$

$$og_t = w_5 h_t + w_6 h'_t. \tag{16}$$

4. Performance validation

In this section, the performance outcomes of the ASLA-DLHGR methodology are examined and tested using the ASL dataset [31]. The technique is simulated using Python 3.6.5 on a PC with an i5-8600k, 250GB SSD, GeForce 1050Ti 4GB, 16GB RAM, and 1TB HDD. The parameters include a learning rate of 0.01, ReLU activation, 50 epochs, 0.5 dropout, and a batch size of 5. Figure 3 represents the sample images.

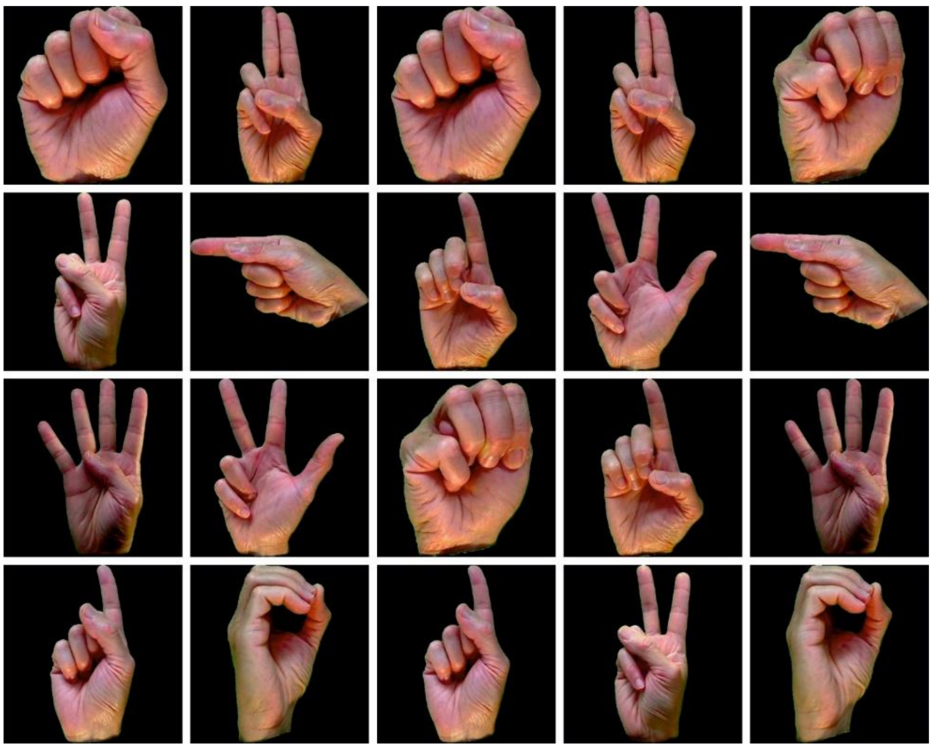

Figure 3. Sample images.

In Table 1 and Figure 4, the overall SL recognition outcome of the ASLA-DLHGR model is described in terms of definite aspects. The outcomes emphasized that the ASLA-DLHGR model effectively identify several kinds of signs. Furthermore, it is seen that the ASLA-DLHGR approach gains average $prec_n$, $reca_l$, $accu_y$, and F_{score} of 99.97%, 99.97%, 99.96%, and 99.98%, respectively.

Table 1. Overall SL detection of the ASLA-DLHGR approach under distinct measures.

Sign	$Prec_n$	$Reca_l$	$Accu_y$	F_{Score}	Sign	$Prec_n$	$Reca_l$	$Accu_y$	F_{Score}
0	99.92	99.98	99.98	99.99	I	100.00	99.99	99.90	100.00
1	100.00	100.00	99.99	99.98	J	100.00	99.99	100.00	99.99
2	100.00	99.91	100.00	99.98	K	100.00	99.99	99.99	99.99
3	99.94	99.91	99.91	99.98	L	99.97	99.99	99.97	99.98
4	100.00	99.90	99.97	99.98	M	99.98	99.96	99.98	99.99
5	99.94	99.99	100.00	99.98	N	100.00	100.00	99.98	99.99
6	100.00	99.91	99.94	99.99	O	100.00	99.99	100.00	99.99
7	99.99	100.00	100.00	99.99	P	99.94	99.98	99.91	99.99
8	100.00	99.99	99.95	99.98	Q	100.00	100.00	100.00	99.98
9	99.99	99.99	99.97	99.99	R	100.00	99.99	99.99	99.98
A	100.00	99.90	99.94	99.99	S	99.95	99.94	100.00	99.98
B	99.91	99.99	99.90	99.99	T	100.00	100.00	99.99	99.99
C	99.93	99.99	100.00	99.99	U	100.00	99.99	100.00	99.99
D	99.91	99.98	99.98	99.99	V	100.00	99.99	99.99	99.99
E	99.93	99.98	100.00	99.99	W	99.99	100.00	99.98	99.99
F	100.00	99.98	99.98	99.99	X	100.00	99.99	99.98	99.99
G	99.93	99.98	99.99	99.98	Y	99.98	99.96	100.00	100.00
H	99.91	99.98	100.00	99.98	Z	99.99	99.99	99.94	99.99
					Average	99.98	99.98	99.98	99.99

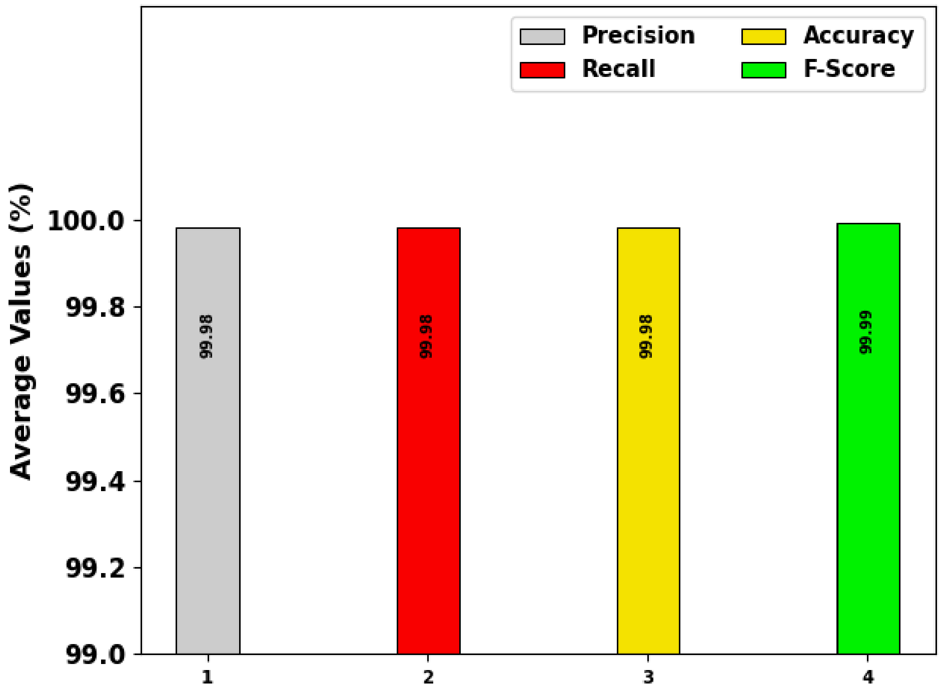

Figure 4. Average results of the ASLA-DLHGR approach under distinct measures.

In Figure 5, the TRA $accu_y$ (TRAAY) and validation $accu_y$ (VLAAY) curves of the ASLA-DLHGR approach are exhibited. The $accu_y$ values are computed over a range of 0-25 epochs. The outcome highlights that the TRAAY and VLAAY analysis displays a rising tendency, which informed the capacity of the ASLA-DLHGR approach with superior performance over many iterations. Besides, the TRAAY and VLAAY remain closer over the epochs that specify lesser overfitting and show higher performance of the ASLA-DLHGR model, guaranteeing consistent prediction on hidden instances.

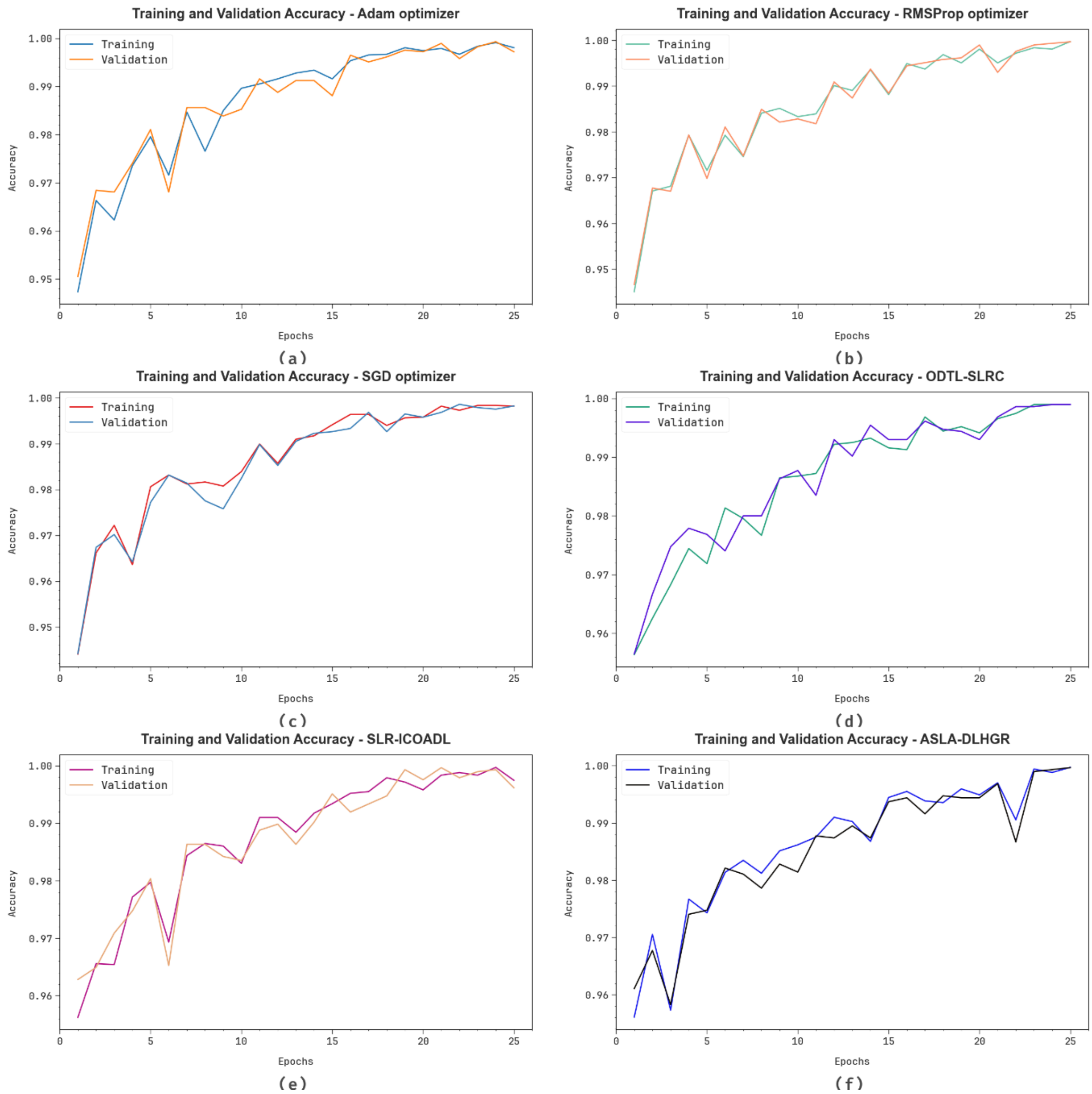

Figure 5. $Accu_y$ curve of the ASLA-DLHGR method.

In Figure 6, the TRA loss (TRALO) and VLA loss (VLALO) curve of the ASLA-DLHGR technique is demonstrated. The loss outcomes are computed over the range of 0-25 epochs. It is indicated that the TRALO and VLALO values exemplify a reducing tendency, which highlights the capacity of the ASLA-DLHGR methodology in balancing a trade-off between data fitting and generalization. Additionally, the constant decrease in loss outcomes ensures the maximal effectiveness of the ASLA-DLHGR methodology and tunes the prediction solution with time.

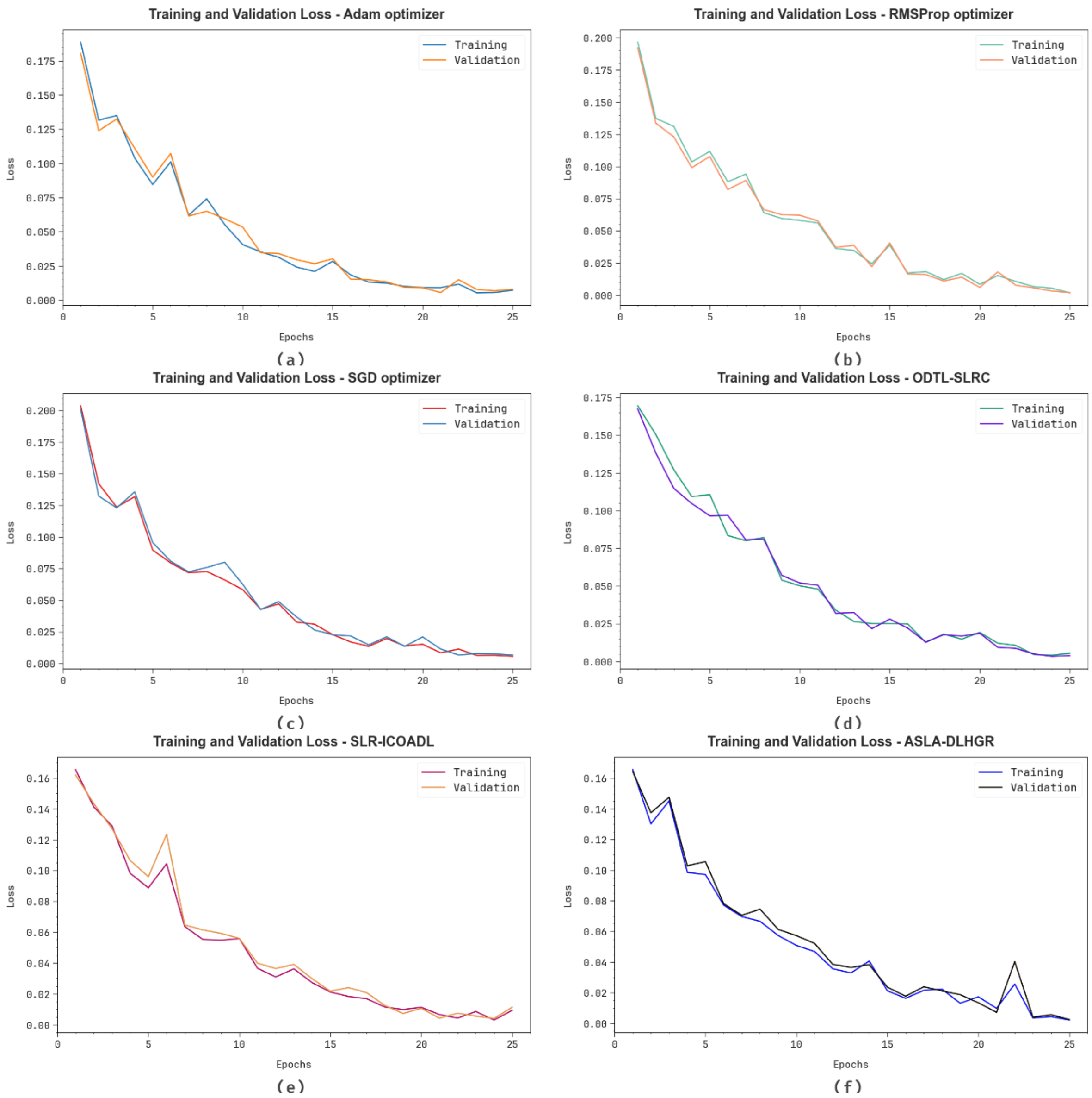

Figure 6. Loss graph of the ASLA-DLHGR method.

To demonstrate the proficiency of the ASLA-DLHGR technique, a comprehensive comparison of outcomes is presented in Table 2 [32].

In Figure 7, the relative $prec_{ision}$ and $reca_{ll}$ outcomes of the ASLA-DLHGR model are presented. The outcomes indicate that the Adam optimizer method displays worse values with $prec_n$ and $reca_l$ of 99.86% and 99.81%. In the meantime, the SGD optimizer and RMSProp optimizer approach achieve slightly enhanced $prec_n$ and $reca_l$. Concurrently, the ODTL-SLRC and SLR-ICOADL methodologies establish closer values of $prec_n$ and $reca_l$. However, the ASLA-DLHGR approach provides superior performance with $prec_n$ and $reca_l$ of 99.98% and 99.98%, respectively.

Table 2. Comparative examination of ASLA-DLHGR methodology with existing models.

Methods	$Prec_n$	$Reca_l$	$Accu_y$	F_{Score}
ASLA-DLHGR	99.98	99.98	99.98	99.99
SLR-ICOADL	99.96	99.96	99.95	99.96
ODTL-SLRC	99.90	99.90	99.91	99.92
SGD	99.87	99.87	99.87	99.82
RMSProp	99.89	99.86	99.72	99.84
Adam	99.86	99.81	99.22	99.87

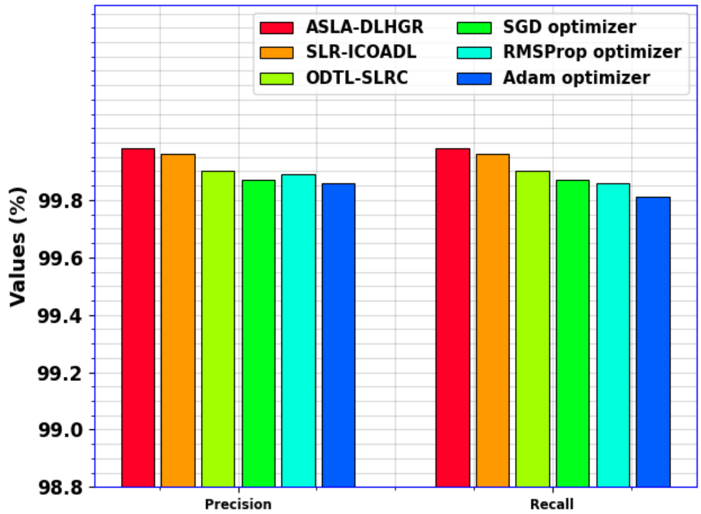

Figure 7. $Prec_n$ and $Reca_l$ results of ASLA-DLHGR methodology with existing models.

In Figure 8, a comparative $accu_y$ and F_{score} result of the ASLA-DLHGR methodology is provided. The outcomes indicate that the Adam optimizer system provides worse values, with $accu_y$ and F_{score} of 99.22% and 99.87%. At the same time, the SGD optimizer and RMSProp optimizer methods have slightly better $accu_y$ and F_{score}. Meanwhile, the ODTL-SLRC and SLR-ICOADL methodologies depict closer values of $accu_y$ and F_{score}. Nevertheless, the ASLA-DLHGR model results in enhanced performance with $accu_y$ and F_{score} of 99.98% and 99.99%.

The recognition rate (RR), computation time (CT), and results of the ASLA-DLHGR model are related to other existing methodologies in Table 3. Figure 9 exhibits the comparative RR outcome of the ASLA-DLHGR technique. The results signify that the ASLA-DLHGR technique gains a maximum RR of 99.97%. On the other hand, the KNN, SVM, ANN, CNN, ODTL-SLRC, and SLR-ICOADL methods attain minimal RR values of 96.23%, 98.07%, 98.08%, 99.86%, 99.90%, and 99.92%, respectively.

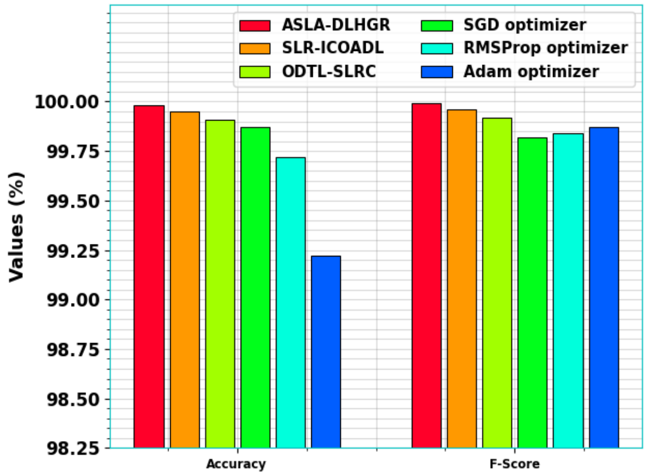

Figure 8. $Accu_y$ and F_{score} outcomes of ASLA-DLHGR methodology with existing models.

Table 3. RR and CT outcomes of the ASLA-DLHGR model with existing approaches.

Methods	RR (%)	CT (min)
KNN	96.23	16.56
SVM Classifier	98.07	14.37
ANN Method	98.08	15.44
CNN Technique	99.86	11.23
ODTL-SLRC	99.90	6.46
SLR-ICOADL	99.92	3.93
ASLA-DLHGR	99.97	1.34

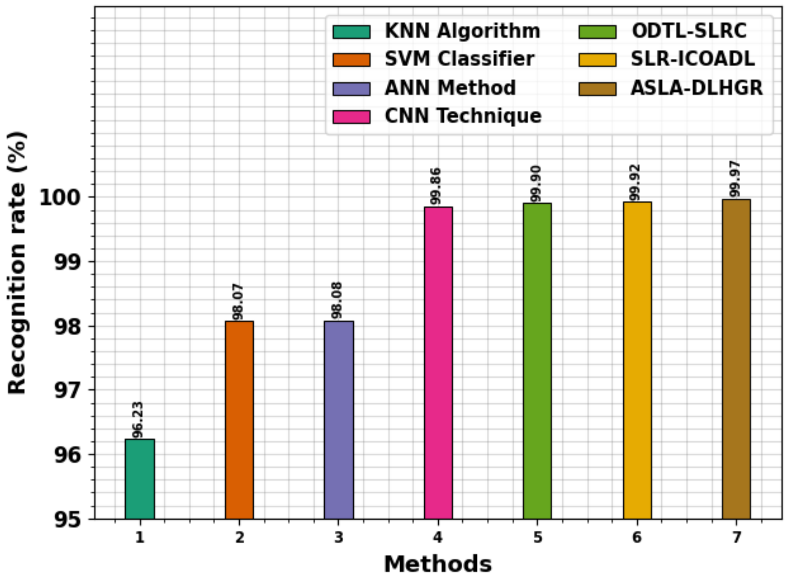

Figure 9. RR outcome of the ASLA-DLHGR model with existing approaches.

Figure 10 illustrates the comparative CT analysis of the ASLA-DLHGR approach. The outcomes represented that the ASLA-DLHGR approach reaches a lesser CT of 1.34min. Ultimately, the KNN, SVM, ANN, CNN, ODTL-SLRC, and SLR-ICOADL methodologies yield better CT values of 16.56min, 14.37min, 15.44min, 11.23min, 6.46min, and 3.93min, respectively.

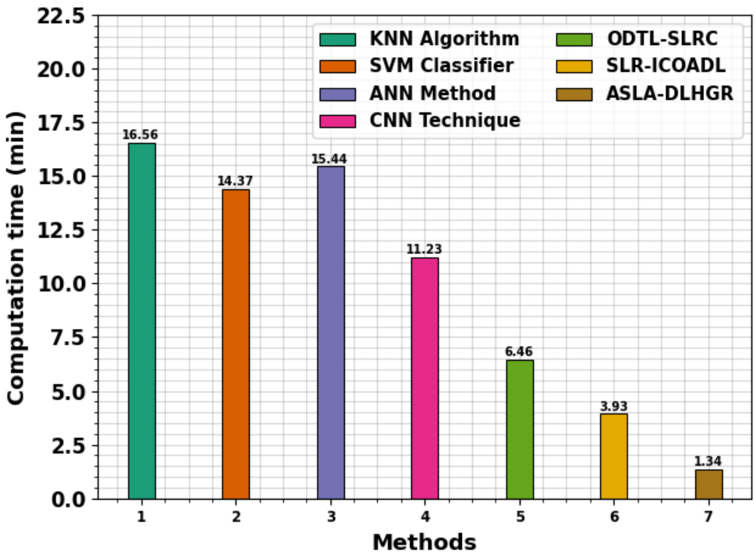

Figure 10. CT outcome of the ASLA-DLHGR model with existing approaches.

Table 4 and Figure 11 depict the error analyses of the ASLA-DLHGR method with existing models. The ASLA-DLHGR method exhibits poor performance with an $accu_y$ of 2%, $prec_n$ of 2%, $reca_l$ of 2%, and F_{score} of 1%. The SLR-ICOADL technique illustrates a slight improvement with an $accu_y$ of 5%, $prec_n$ of 4%, $reca_l$ of 4%, and F_{score} of 4%, though still ineffective. ODTL-SLRC performs better with an $accu_y$ of 9%, $prec_n$ of 10%, $reca_l$ of 10%, and F_{score} of 8%, yet remains limited in overall reliability. SGD achieves a balanced performance with an $accu_y$, $prec_n$, and $reca_l$ at 13%, and the highest F_{score} of 18% among conventional methods. RMSProp provides an improved $accu_y$ of 28%, with $prec_n$ of 11%, $reca_l$ of 14%, and F_{score} of 16%, depicting moderate learning ability. Adam optimizer demonstrates the best $accu_y$ at 78%, along with $prec_n$ of 14%, $reca_l$ of 19%, and F_{score} of 13%, which suggests that effectively captures the classification patterns despite slightly lower $prec_n$ and F_{score}, emphasizing class imbalance or overfitting.

Table 4. Error analysis of the ASLA-DLHGR method with existing models.

Methods	$Prec_n$	$Reca_l$	$Accu_y$	F_{Score}
ASLA-DLHGR	0.02	0.02	0.02	0.01
SLR-ICOADL	0.04	0.04	0.05	0.04
ODTL-SLRC	0.10	0.10	0.09	0.08
SGD	0.13	0.13	0.13	0.18
RMSProp	0.11	0.14	0.28	0.16
Adam	0.14	0.19	0.78	0.13

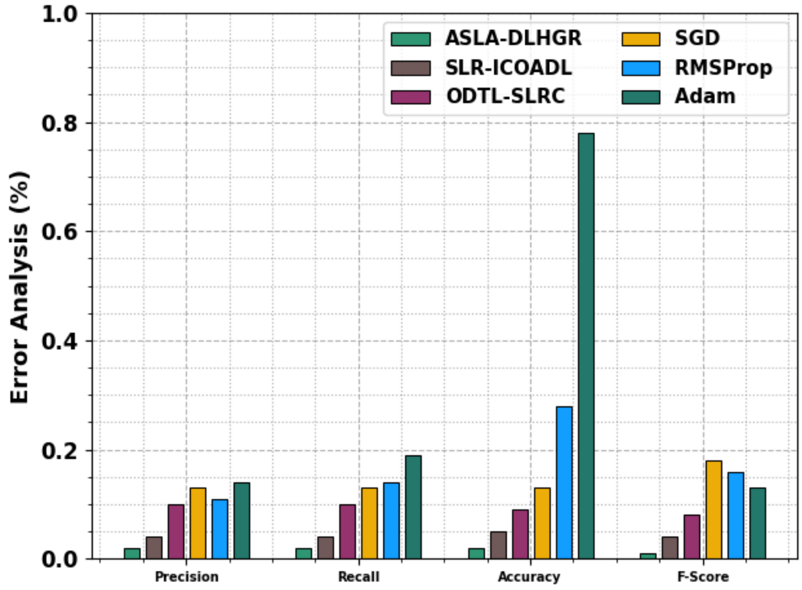

Figure 11. Error analysis of the ASLA-DLHGR method with existing models.

Table 5 demonstrates the ablation study analysis of the ASLA-DLHGR technique with existing methods. The ASLA-DLHGR technique demonstrates the highest performance with an $accu_y$ of 99.98%, $prec_n$ of 99.98%, $reca_l$ of 99.98%, and F_{score} of 99.99%. CNN-BiLSTM+TSA with parameter tuning but without feature extraction achieves an $accu_y$ of 99.22%, $prec_n$ of 99.27%, $reca_l$ of 99.36%, and F_{score} of 99.22%, highlighting robust performance but slightly lower than the fully optimized method. CNN-BiLSTM+SqueezeNet with feature extraction but without parameter tuning provides an $accu_y$ of 98.68%, $prec_n$ of 98.72%, $reca_l$ of 98.74%, and F_{score} of 98.43%, which suggests that tuning significantly enhances performance even when features are well extracted. The base CNN-BiLSTM model without feature extraction or tuning results in an $accu_y$ of 98.07%, $prec_n$ of 98.02%, $reca_l$ of 98.09%, and F_{score} of 97.79%, highlighting the benefit each component contributes in the overall process.

Table 5. Comparative performance evaluation of the ASLA-DLHGR technique through ablation study against existing methods.

Techniques	$Prec_n$	$Reca_l$	$Accu_y$	F_{Score}
ASLA-DLHGR (with feature extraction and parameter tuning)	99.98	99.98	99.98	99.99
CNN-BiLSTM+TSA (without feature extraction with parameter tuning)	99.27	99.36	99.22	99.22
CNN-BiLSTM+SqueezeNet (with feature extraction without parameter tuning)	98.72	98.74	98.68	98.43
CNN-BiLSTM	98.02	98.09	98.07	97.79

5. Conclusions

In this article, the ASLA-DLHGR technique is proposed for the hearing and speech-impaired.

The goal of the ASLA-DLHGR technique is to recognize hand gestures to communicate among disabled people. To accomplish that, the ASLA-DLHGR approach comprises four processes, pre-processing, feature extraction, parameter choice, and classification process. Initially, the ASLA-DLHGR approach performs data pre-processing using BF. Furthermore, the ASLA-DLHGR technique implements the SqueezeNet model to learn composite features from the pre-processed data. To develop the performance of the SqueezeNet technique, the TSA-based hyperparameter process is performed. For the gesture recognition process, the hybrid of the CNN-BiLSTM method is employed. The experimental validation of the ASLA-DLHGR method demonstrated a superior accuracy value of 99.98% over existing models under the ASL dataset. The limitations include a dependence on controlled datasets that may not fully capture the variability of real-world conditions, such as diverse backgrounds, lighting, and occlusions. Moreover, the efficiency of the model across diverse cultural SLs and individual discrepancies requires sufficient testing. The deployment on low-power or portable devices may be restricted by the computational requirements of the model. It is recommended to supplement the study with more data and case studies regarding the number of hearing and visually impaired individuals globally or in specific regions, as well as the current status of the use of communication aids. Future work may concentrate on expanding data diversity, improving system adaptability to diverse environments, and exploring user-centric evaluations to enhance practical applicability.

Author contributions

Najm Alotaibi: Conceptualization, methodology, validation, investigation, writing-original draft preparation; Alanoud Subahi and Nouf Atiahallah Alghanmi: Conceptualization, methodology, writing-original draft preparation, writing-review and editing; Mohammed Rizwanullah: Software, data creation, visualization, validation, writing-review and editing. All authors have read and agreed to the published version of the manuscript.

Use of Generative-AI tools declaration

The authors declare that they have not used Artificial Intelligence (AI) tools in the creation of this article.

Data availability statement

The data that support the findings of this study are openly available in the Kaggle repository at https://www.kaggle.com/datasets/ayuraj/asl-dataset, reference number [32].

Acknowledgements

The authors extend their appreciation to the King Salman Centre for Disability Research for funding this work through Research Group No. KSRG-2024-343.

Conflict of interest

The authors declare no conflicts of interest.

References

1. M. M. Asiri, A. Motwakel, S. Drar, Enhanced bald eagle search optimizer with transfer learning-based sign language recognition for hearing-impaired persons, *J. Disabil. Res.*, **2** (2023), 86–93. https://doi.org/10.57197/JDR-2023-0039
2. F. Alrowais, S. S. Alotaibi, S. Dhahbi, R. Marzouk, A. Mohamed, A. M. Hilal, Sign language recognition and classification model to enhance quality of disabled people, CMC-*Comput. Mater. Con.*, **73** (2022), 3419–3432. https://doi.org/10.32604/cmc.2022.029438
3. M. Zakariah, Y. A. Alotaibi, D. Koundal, Y. H. Guo, M. Mamun Elahi, Sign language recognition for Arabic alphabets using transfer learning technique, *Comput. Intel. Neurosc.*, **2022** (2022), 4567989. https://doi.org/10.1155/2022/4567989
4. N. K. B. Duraimutharasan, K. Sangeetha, Machine learning and vision based techniques for detecting and recognizing Indian sign language, *Revue d'Intelligence Artificielle*, **37** (2023), 1361–1366. https://doi.org/10.18280/ria.370529
5. M. Potnis, D. Raul, M. Inamdar, Recognition of Indian sign language using machine learning algorithms, *2021 8th International Conference on Signal Processing and Integrated Networks (SPIN)*, Noida, India, 2021, 579–584. https://doi.org/10.1109/SPIN52536.2021.9566141
6. E. Daniel, V. Kathiresan, C. Priyadarshini, R. G. Nancy P. Sindhu, Real time sign recognition using YOLOv8 object detection algorithm for Malayalam sign language, *Fusion: Practice and Applications*, **17** (2025), 135–145. https://doi.org/10.54216/FPA.170110
7. A. A. Alhussan, M. M. Eid, W. H. Lim, Advancing communication for the deaf: A convolutional model for Arabic sign Language recognition, *Journal of Artificial Intelligence and Metaheuristics*, **5** (2023), 38–45. https://doi.org/10.54216/JAIM.050104
8. D. Kothadiya, C. Bhatt, K. Sapariya, K. Patel, A.-B. Gil-González, J. M. Corchado, Deepsign: Sign language detection and recognition using deep learning, *Electronics*, **11** (2022), 1780. https://doi.org/10.3390/electronics11111780
9. B. T. Abeje, A. O. Salau, A. D. Mengistu, N. K. Tamiru, Ethiopian sign language recognition using deep convolutional neural network, *Multimed. Tools Appl.*, **81** (2022), 29027–29043. https://doi.org/10.1007/s11042-022-12768-5
10. R. S. A. Ameer, M. A. Ahmed, Z. T. Al-Qaysi, M. M. Salih, M. L. Shuwandy, Empowering communication: A deep learning framework for arabic sign language recognition with an attention mechanism, *Computers*, **13** (2024), 153. https://doi.org/10.3390/computers13060153
11. A. Pandey, A. Chauhan, A. Gupta, V. Karnatak, Voice based Sign Language detection for dumb people communication using machine learning, *J. Pharm. Negat. Result.*, **14** (2023), 22–30.
12. R. Valarmathi, P. J. Surya, P. Balaji, K. Ashik, Animated sign language for people with speaking and hearing disability using deep learning, *2024 International Conference on Communication, Computing and Internet of Things (IC3IoT)*, Chennai, India, 2024, 1–5. http://doi.org/10.1109/ic3iot60841.2024.10550211
13. M. Jebali, A. Dakhli, W. Bakari, Deep learning-based sign language recognition system for cognitive development, *Cogn. Comput.*, **15** (2023), 2189–2201. https://doi.org/10.1007/s12559-023-10182-z

14. A. S. M. Miah, M. Al M. S. Nishimura, J. Shin, Sign language recognition using graph and general deep neural network based on large scale dataset, *IEEE Access*, **12** (2024), 34553–34569. https://doi.org/10.1109/ACCESS.2024.3372425
15. V. Bhatt, R. Dash, Real-time hand gesture recognition for American sign language using CNN, mediapipe and convexity approach, In: *Machine Learning, Image Processing, Network Security and Data Sciences*, Cham: Springer, 2024, 260–271. https://doi.org/10.1007/978-3-031-62217-5_22
16. J. Shin, A. S. M. Miah, Y. Akiba, K. Hirooka, N. Hassan, Y. S. Hwang, Korean sign language Alphabet recognition through the integration of handcrafted and deep learning-based two-stream feature extraction approach, *IEEE Access*, **12** (2024), 68303–68318. https://doi.org/10.1109/ACCESS.2024.3399839
17. B. B. Jayasingh, K. M. S. Rani, K. Swathi, Hand gestures translation system for sgn language using CNN models, *2025 4th OPJU International Technology Conference (OTCON) on Smart Computing for Innovation and Advancement in Industry 5.0*, Raigarh, India, 2025, 1–6. https://doi.org/10.1109/OTCON65728.2025.11071065
18. K. Aurangzeb, K. Javeed, M. Alhussein, I. Rida, S. I. Haider, A. Parashar, Deep learning approach for hand gesture recognition: applications in deaf communication and healthcare, CMC-*Comput. Mater. Con.*, **78** (2024), 127–144. https://doi.org/10.32604/cmc.2023.042886
19. A. T. Elgohr, M. S. Elhadidy, M. El-geneedy, S. Akram, M. A. A. Mousa, Advancing sign language recognition: A YOLO v. 11-based deep learning framework for Alphabet and transactional hand gesture detection, *Proceedings of the AAAI Symposium Series*., **6** (2025), 209–217. https://doi.org/10.1609/aaaiss.v6i1.36055
20. C. Rathnayake, R. Gamage, K. D. R. N. Kalubowila, S. Thilakarathne, Development of a real-time Hand gesture recognition system for aid of hearing-impaired communication using flex sensors and machine learning algorithms, In: *2024 8th SLAAI International Conference on Artificial Intelligence (SLAAI-ICAI)*, 2024, 1–6. https://doi.org/10.1109/SLAAI-ICAI63667.2024.10844963
21. S. Malviya, A. Mahajan, K. K. Sethi, Machine learning framework for intelligent hand gesture recognition: An application to Indian sign language and hand talk, In: *Proceedings of the International Conference on Recent Advancements and Modernisations in Sustainable Intelligent Technologies and Applications (RAMSITA 2025)*, Paris: Atlantis Press, 2025, 496–512. https://doi.org/10.2991/978-94-6463-716-8_39
22. A. S. M. Miah, M. A. M. Hasan, Y. Tomioka, J. Shin, Hand gesture recognition for multi-culture sign language using graph and general deep learning network, *IEEE Open Journal of the Computer Society*, **5** (2024), 144–155. https://doi.org/10.1109/OJCS.2024.3370971
23. P. Singhal, S. Verma, R. Gupta, R. Kumar, R. K. Arya, Vision-based hand gesture recognition system for assistive communication using neural networks and GSM integration, *2025 2nd International Conference on Computational Intelligence, Communication Technology and Networking (CICTN)*, Ghaziabad, India, 2025, 891–895. https://doi.org/10.1109/CICTN64563.2025.10932562
24. A. Rehman, M. Zaman, T. Kehkashan, F. Akbar, M. Hamza, R. A. Riaz, Enhanced sign language detection with deep CNN: Achieving accuracy in hand gesture recognition, *2024 5th International Conference on Innovative Computing (ICIC)*, Lahore, Pakistan, 2024, 1–6. https://doi.org/10.1109/ICIC63915.2024.11116573

25. C. M. Soukaina, M. Mohammed, R. Mohamed, Geometric feature-based machine learning for efficient hand sign gesture recognition, *Statistics, Optimization & Information Computing*, **13** (2025), 2027–2043. https://doi.org/10.19139/soic-2310-5070-2306
26. S. Kukreja, A. Singh, V. Chauhan, Hand gesture vision: Integrating computer vision and machine learning for enhanced communication, *2024 1st International Conference on Advances in Computing, Communication and Networking (ICAC2N)*, Greater Noida, India, 2024, 35–40. https://doi.org/10.1109/ICAC2N63387.2024.10895070
27. L. S. Wu, L. Y. Fang, J. Yue, B. Zhang, P. Ghamisi, M. He, Deep bilateral filtering network for point-supervised semantic segmentation in remote sensing images, *IEEE T. Image Process.*, **31** (2022), 7419–7434. https://doi.org/10.1109/TIP.2022.3222904
28. I. Camelo, Convolutional neural network-based object detection with limited embedded computational resources, PhD Thesis, Université du Québec en Outaouais, 2024.
29. V. Chandran, P. Mohapatra, A novel reinforcement learning-inspired tunicate swarm algorithm for solving global optimization and engineering design problems, *J. Ind. Manag. Optim.*, **21** (2025), 565–612. https://doi.org/10.3934/jimo.2024095
30. J. Li, Y. X. Wang, W. S. Liang, C. Xiong, W. B. Cai, L. J. Li, et al., Visual anomaly detection via CNN-BiLSTM network with knit feature sequence for Floating-Yarn stacking during the high-speed sweater knitting process, *Electronics*, **13** (2024), 3968. https://doi.org/10.3390/electronics13193968
31. A. Thakur, *American sign language dataset*, 2019. Available from: https://www.kaggle.com/datasets/ayuraj/asl-dataset.
32. M. M. Asiri, A. Motwakel, S. Drar, Robust sign language detection for hearing disabled persons by improved coyote optimization algorithm with deep learning, *AIMS Mathematics*, **9** (2024), 15911–15927. https://doi.org/10.3934/math.2024769