



OPEN

IoT-driven smart assistive communication system for the hearing impaired with hybrid deep learning models for sign language recognition

Mashael Maashi¹✉, Huda G. Iskandar^{2,3} & Mohammed Rizwanullah⁴

Deaf and hard-of-hearing people utilize sign language recognition (SLR) to interconnect. Sign language (SL) is vital for hard-of-hearing and deaf individuals to communicate. SL uses varied hand gestures to speak words, sentences, or letters. It aids in linking the gap of communication between individuals with hearing loss and other persons. Also, it creates comfortable for individuals with hearing loss to convey their feelings. The Internet of Things (IoTs) can help persons with disabilities sustain their desire to attain a good quality of life and permit them to contribute to their economic and social lives. Modern machine learning (ML) and computer vision (CV) developments have allowed SL gesture detection and decipherment. This study presents a Smart Assistive Communication System for the Hearing-Impaired using Sign Language Recognition with Hybrid Deep Learning (SACHI-SLRHDL) methodology in IoT. The SACHI-SLRHDL technique aims to assist people with hearing impairments by creating an intelligent solution. At the primary stage, the SACHI-SLRHDL technique utilizes bilateral filtering (BF) for image pre-processing to increase the excellence of the captured images by reducing noise while preserving edges. Furthermore, the improved MobileNetV3 model is employed for the feature extraction process. Moreover, the convolutional neural network with a bidirectional gated recurrent unit and attention (CNN-BiGRU-A) model classifier is implemented for the SLR process. Finally, the attraction-repulsion optimization algorithm (AROA) adjusts the hyperparameter values of the CNN-BiGRU-A method optimally, resulting in more excellent classification performance. To exhibit the more significant solution of the SACHI-SLRHDL method, a comprehensive experimental analysis is performed under an Indian SL dataset. The experimental validation of the SACHI-SLRHDL method portrayed a superior accuracy value of 99.19% over existing techniques.

Keywords Sign Language Recognition, Hybrid deep learning, Communication systems, Hearing impaired people, MobileNetV3

Deaf is a disability that impairs their hearing and makes them not able to hear, while dumb is a disability which impairs their speaking and makes them not able to speak¹. Both are only disabled in speaking and hearing; they can still do many other things. The only thing that separates them and the normal people interaction. If there is a way for normal people and deaf-dumb individuals to interact, the deaf-dumb individual can live like normal people². And the only way for them to interact is through SL. SL is the primary communication method among hearing-impaired individuals and other people. It is stated through the manual that hand and body motions and non-manual motions are features of facial expression. These characteristics are associated with forming utterances, which transfer the meaning of sentences or words³. SL technology covers a wide-ranging spectrum, from taking signs to their realistic demonstration to assist the interaction among hearing-impaired individuals and between speaking and hearing-impaired individuals. More particularly, SL capturing includes

¹Department of Software Engineering, College of Computer and Information Sciences, King Saud University, PO Box 103786, 11543 Riyadh, Saudi Arabia. ²Department of Information Systems, Faculty of Computer and Information Technology, Sana'a University, Sana'a, Yemen. ³King Salman Center for Disability Research, 11614 Riyadh, Saudi Arabia. ⁴Department of Computer and Self Development, Preparatory Year Deanship, Prince Sattam bin Abdulaziz University, AlKharj, Saudi Arabia. ✉email: mmaashi@ksu.edu.sa

precisely extracting mouth, body, and hand expressions utilizing suitable sensing gadgets in marker-based or marker-less setups⁴.

Accurately capturing the SL technology is presently limited by the discrimination and resolution for sensor capability and the reality that blocks and fast hand movements pose substantial tasks to precisely capturing signs. SLR includes enlarging prevailing ML models to firmly categorize human articulations into continuous sentences or isolated signs⁵. Existing challenges in SLR exist in the absence of massive annotated databases, which significantly affect the precision and generalization capability of SLR approaches and the complexity of identifying sign limitations in continuous SLR consequences. An automated SLR method can identify sign gestures. The gestures are typically given with hand movement and supportive aspects of body postures and facial expressions⁶. SLR contains the entire process of identifying and tracking the signs achieved and changing them into semantically meaningful expressions and words⁷. The SLR method utilizing the IoT and Deep Learning (DL) has been advanced⁸. It is a set of models and techniques with higher-level abstractions through structures formed by several non-linear transformations. DL models employ many data to extract features automatically, aiming to emulate the human brain's capability to observe, analyze, learn, and make an implication, particularly for enormous, diverse complexities⁹. DL structures generate relationships beyond instant neighbours in data and create learning patterns, removing representations directly from data without human intervention¹⁰. Various DL methods have recently been presented for hand gesture detection. A deep convolutional neural network (DCNN) based static SLR.

This study presents a Smart Assistive Communication System for the Hearing-Impaired using Sign Language Recognition with Hybrid Deep Learning (SACHI-SLRHDL) methodology in IoT. The SACHI-SLRHDL technique aims to assist people with hearing impairments by creating an intelligent solution. At the primary stage, the SACHI-SLRHDL technique utilizes bilateral filtering (BF) for image pre-processing to increase the excellence of the captured images by reducing noise while preserving edges. Furthermore, the improved MobileNetV3 model is employed for the feature extraction process. Moreover, the convolutional neural network with a bidirectional gated recurrent unit and attention (CNN-BiGRU-A) model classifier is implemented for the SLR process. Finally, the attraction-repulsion optimization algorithm (AROA) adjusts the hyperparameter values of the CNN-BiGRU-A method optimally, resulting in more excellent classification performance. To exhibit the more significant solution of the SACHI-SLRHDL method, a comprehensive experimental analysis is performed under an Indian SL dataset. The key contribution of the SACHI-SLRHDL method is listed below.

- The SACHI-SLRHDL model utilizes BF for pre-processing to improve image quality by effectively reducing noise while preserving edges. This enhances the clarity of input images, which is significant for accurate SLRs. The model ensures improved performance in subsequent stages of the recognition process by improving image features.
- The SACHI-SLRHDL approach uses an improved MobileNetV3 architecture for feature extraction, optimizing performance and computational efficiency. This adaptation enables faster processing without compromising the accuracy of extracted features. Mitigating complexity makes the model more appropriate for real-time applications in SLR.
- The SACHI-SLRHDL methodology introduces a novel CNN-BiGRU-A framework that integrates CNN for spatial feature extraction and BiGRU for processing sequential data. Attention mechanisms (AMs) are integrated to prioritize critical features, improving the method's capability to recognize intrinsic SL gestures. This approach improves both the accuracy and interpretability of the recognition system.
- The SACHI-SLRHDL approach employs the AROA model for fine-tuning, improving the training process by effectively optimizing the model's parameters. This method enhances overall performance by balancing exploration and exploitation during optimization. The role of AROA is crucial in achieving greater efficiency and accuracy in the recognition task.
- The key novelty of the SACHI-SLRHDL approach is the integration of CNN, BiGRU, AMs, and the AROA model to improve SLR. This hybrid model enables effective spatial and temporal feature extraction while the AM assists in prioritizing critical features. Additionally, AROA optimizes the model's performance, ensuring higher accuracy and efficiency in real-time applications.

Literature works

Akhila Thejaswi et al.¹¹ investigate a combination of SL Recognition (SLR) and SL Translation (SLT) methods to guarantee accurate real-world sign gesture recognition. A CNN is utilized in the presented work as a DL method to train a more extensive database of hand gestures and achieve image investigation employing the MediaPipe library for identification and landmark estimation. A vast dataset analysis organizes the study. In¹², an IoT-based method, which can fit on a ring finger, is presented. This method can translate and learn English and Arabic braille into audio utilizing DL methods improved with transfer learning (TL). The detection of the braille image captured is attained through a TL-based CNN. Shwany et al.¹³ project a real-world approach to detect threatening signs by criminals through interrogation. The presented approach is to install a proper camera feature in front of the offender, hand gesture records in a particular area of the hand, employ a few image processing methods, like contrast improvement methods, to the image to help detection as input, and then categorize the image utilizing CNN for a particular concern with enhanced capabilities for that area and utilizing AlexNet. Lakshmi et al.¹⁴ intend to generate a real-world video-based interactive SLT education method. The Flask framework and pre-trained methods for classification are segments of the method to enhance communication and understanding with SL. In the future, the presented method will utilize Python libraries and tensor flow for image input forecasting, which can efficiently resolve several interaction complexities. The presented method represents an engaging and innovative method to boost and teach communication capabilities in deaf and dumb children. Akdag and Baykan¹⁵ present a novel method to isolate SL word recognition utilizing an innovative

DL method, which associates the assets both temporally separated (R(2+1)D) and residual three-dimensional (R3D) convolutional blocks. The R3(2+1)D-SLR method can take the complicated temporal and spatial features vital for precise sign detection.

In¹⁶, an ensemble meta-learning method is projected. The work tests and trains the deep ensemble meta-learning method utilizing dual synthetically created assistive service databases. The DL method utilizes several ensemble input learners to employ a meta-classification system shared with each output, demonstrating individual assistive services. This method attains substantially greater outcomes than classical ML methods and simpler feed-forward neural network methods without the ensemble method. Parveen et al.¹⁷ present a gadget that can change a deaf individual hand gestures into voice and text. The video recorded by the camera employs an open-source CV toolkit named OpenCV. Then, the video utilizes image processing methods, including histogram of gradient (HOG) and CNN. Afterwards, a Raspberry Pi 4 examines the recorded motions and relates the outcomes with a database. Faisal et al.¹⁸ projected three components: (i) the sign recognition module (SRM) that identifies the signs of deaf people, (ii) the speech recognition and synthesis module (SRSM) that procedures the speech of non-deaf people and changes into text, (iii) avatar module (AM) to perform and create the equivalent sign of the non-deaf speech that are incorporated into the sign translation companion method named as SDCS to assist the interaction between deaf to hear and conversely. Li et al.¹⁹ present an updated Archive File Integrity Check Method (AFICM) using a hybrid DL model integrating Bi-directions Long Short-Term Memory (Bi-LSTM) with adaptive gating and Temporal Convolutional Neural Networks (TCNN) model. Ghadi et al.²⁰ explore ML methods used to address security issues in wireless sensor networks while considering their functionality, adaptation challenges, and open problems in the field. Zholshiyeva et al.²¹ explore automating SLT by integrating ML and DL techniques for Kazakh SL (QazSL) recognition. Five algorithms are employed, trained on a dataset of over 4,400 images.

Ghadi et al.²² explore the challenges and safety issues of integrating federated learning (FL) with IoT, focusing on its applications in smart businesses, cities, transportation, and healthcare while addressing encrypted data transmission requirements. Thakur, Dangi, and Lalwani²³ introduce two Hybrid Learning Algorithms (HLA) combining CNN and recurrent neural network (RNN) to capture spatial and sequential patterns, improved by optimization techniques using Whale Optimization and Grey Wolf Optimizer for feature selection. Mazhar et al.²⁴ explore the motivation for IoT device installation in smart buildings and grids, focusing on incorporating artificial intelligence (AI), IoT, and smart grids to improve energy efficiency, security, and comfort while examining ML methods for forecasting energy demand. John and Deshpande²⁵ propose a hybrid deep RNN with Chaos Game Optimization (CGO) for effectual hand gesture recognition, aiming to classify alphabet signs from 2D gesture images through pre-processing, feature extraction, selection, and classification stages. Renjith, Manazhy, and Suresh²⁶ present a hybrid model for Indian SL (ISL) recognition integrating CNNs for spatial feature extraction and RNNs for capturing temporal relationships, aiming to improve the capability of the method to identify complex sign gestures from a dataset of 36 ISL sign classes. Paul et al.²⁷ introduce a novel Human Motion Recognition (HMR) method for medical-related human activities (MRHA) detection, integrating EfficientNet for spatial feature extraction and ConvLSTM for spatio-temporal pattern recognition, followed by a classification module for final predictions. Palanisamy et al.²⁸ present a hybrid approach incorporating DL and graph theory for SLR, illustrating crucial enhancements in accuracy and computational efficiency, making it a competitive solution for enhancing communication for the hearing impaired.

The primary limitation in current research is the lack of sufficiently large and diverse datasets for SLR, which limits the generalizability and accuracy of models across diverse SLs and real-world scenarios. Many existing methods also face difficulty capturing both spatial and temporal features of sign gestures, resulting in mitigated recognition efficiency. Additionally, challenges related to integrating DL and ML methods with IoT, FL, and privacy-preserving techniques must be addressed. Moreover, research on real-time SLT and interaction systems is still in its early stages, with high accuracy and low-latency performance being difficult to attain consistently. Finally, the diversity in regional SL gestures and the dynamic nature of signs pose further hurdles in creating universally applicable systems.

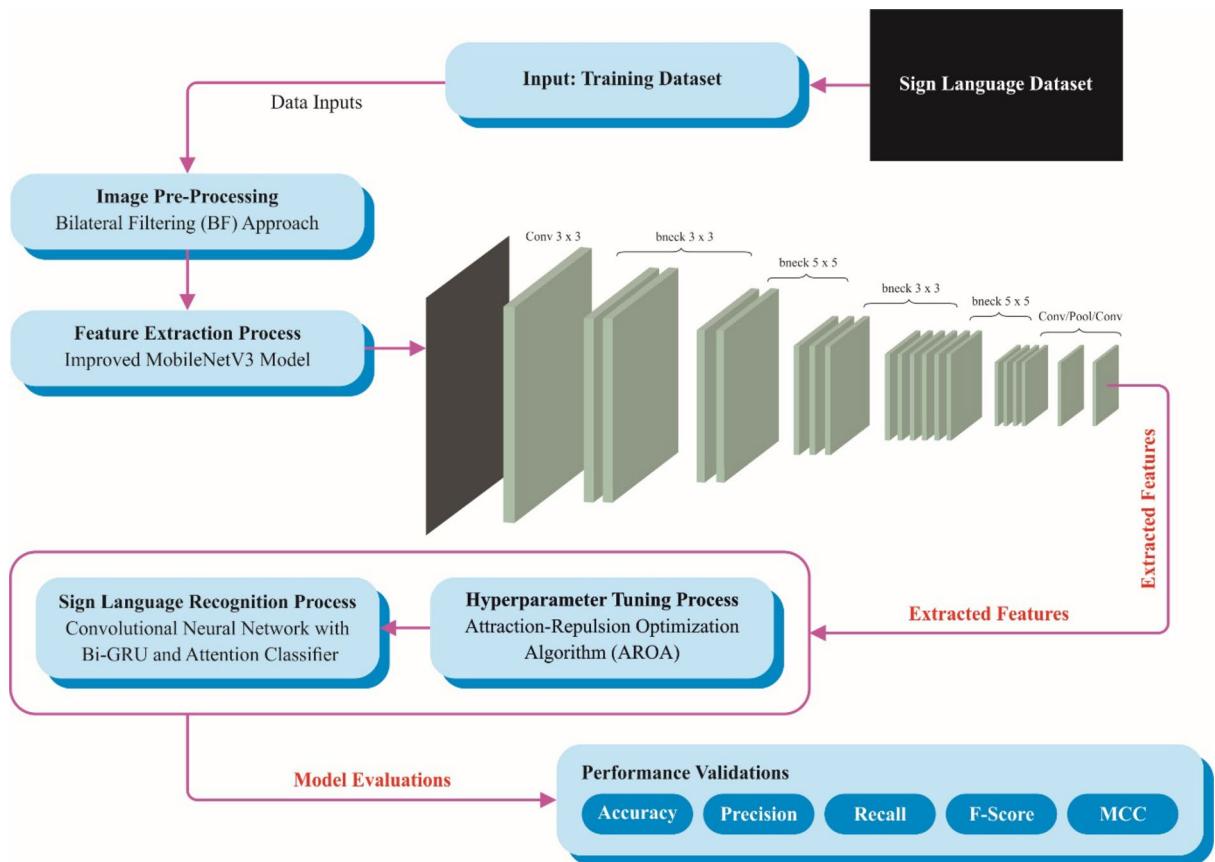
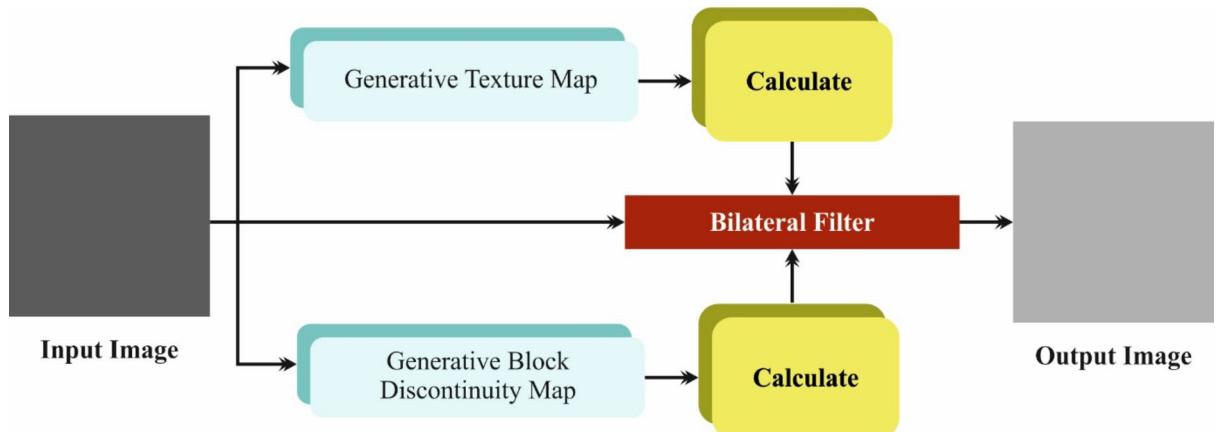
The article is structured as follows: Sect. 2 presents the literature review, Sect. 3 outlines the proposed method, Sect. 4 details the results evaluation, and Sect. 5 concludes the study.

Proposed methodology

This study presents a SACHI-SLRHDL methodology in IoT. The SACHI-SLRHDL technique aims to develop an effective SLR technique that assists people with hearing impairments by creating an intelligent solution. It comprises four distinct processes: image pre-processing, improved MobileNetV3 for feature extractor, hybrid DL classification process, and AROA-based parameter tuning. Figure 1 depicts the entire flow of the SACHI-SLRHDL methodology.

Image pre-processing: BF model

Initially, the SACHI-SLRHDL approach utilizes BF for image pre-processing to improve the excellence of the captured images by decreasing noise while preserving edges²⁹. This model was chosen for image pre-processing due to its superior capability to preserve edges while reducing noise, which is significant for maintaining the integrity of SL gestures. Unlike conventional smoothing techniques, BF effectively smooths out noise without blurring the crucial details, ensuring that key features of the SL images remain intact. Furthermore, BF works well in scenarios with varying lighting conditions and complex backgrounds, which is common in real-world applications. This makes it ideal for pre-processing SL images that may suffer from such challenges. Moreover, the BF model is computationally efficient, allowing it to be implemented in real-time systems, which is significant for SLR tasks. By improving image quality without compromising key spatial details, BF assists in enhancing the overall performance of subsequent DL methods in SLR. Figure 2 specifies the BF architecture.

**Fig. 1.** Overall flow of the SACHI-SLRHDL model.**Fig. 2.** Structure of BF model.

BF is an innovative image pre-processing model that enhances the excellence of images using SLR techniques. It aids in reducing noise while defending edges, which is vital for precisely seizing the hand gestures in SL. In the structure of IoT-based SLR, BF certifies that the seized images from IoT devices, like sensors or cameras, are clear and free from falsification. This pre-processing stage considerably progresses the accuracy of feature extraction by upholding significant spatial details in the imageries. By eliminating unrelated noise, BF permits the detection method to concentrate on profound gestures, safeguarding superior performance in real SLT. Therefore, it contributes to the efficacy of IoT-enabled models in helping persons with hearing loss.

Feature extraction: improved MobileNetV3

Next, the improved MobileNetV3 model extracts relevant features from input images³⁰. This technique was chosen due to its capability to balance high performance with low computational cost, making it appropriate for real-time applications like SLR. Unlike larger networks that require extensive computational resources, MobileNetV3 presents effectual processing without sacrificing accuracy, which is significant for deployment on resource-constrained devices such as IoT systems. Its optimized architecture uses depthwise separable convolutions, which mitigate the number of parameters and computational complexity, making it faster and more efficient. Furthermore, MobileNetV3 performs exceptionally well in extracting discriminative features from images, which is crucial for accurately recognizing SL gestures. By implementing the improved MobileNetV3, the model attains high recognition accuracy while maintaining efficiency, even under varying conditions. This makes it a robust choice compared to conventional, heavier CNN architectures. Figure 3 illustrates the MobileNetV3 model.

This work designated MobileNetV3 from the MobileNet series. The MobileNetV3 method keeps its lightweight features, whereas enduring uses the depth-wise separable convolutional and reversed residual module from the MobileNetV2 method. It improves the bottleneck architecture by combining the Squeeze-and-Excitation (SE) units, reinforcing the importance of significant characteristics and overcoming unimportant ones. Furthermore, the novel hard-swish activation function has been accepted to enhance the system architecture. The MobileNetV3 approach is accessible in larger and smaller versions according to the availability of resources, and this work utilizes the MobileNetV3-larger approach as a base.

However, combining the SE modules into the bottleneck architecture of MobileNetV3-Large has developed the model's performance; the SE modules choose information amongst channels to define the significance of all channels. Nevertheless, it manages the important positioning information within the visual fields. Therefore, this method can only capture local feature information, resulting in problems like scattered fields of interest and narrow performances. To deal with these restrictions, the ECA unit increases the SE modules by preventing the reduction of dimensions and taking cross-channel interaction information more effectively.

Despite developing the ECA unit through the SE modules, it still selects the information amongst channels. During the paper, the SE module is substituted in the MobileNetV3 architecture using the CA module to enhance MobileNetV3. The complete framework of the enhanced MobileNetV3-CA method is presented. The module of CA can concentrate the attention models on the field of interest over the efficient position in the pixel-coordinated method, thus gaining information that considers either position or channel, decreasing the attention to interference data and enhancing the feature appearance capability of the technique. The fundamental architecture of the CA module is presented. For a specified feature graph X , the width is W , the channel counts are C , and the height is H . The module of CA initially pools the input X in dual spatial control, like width and height, to get feature mapping in either direction. Then, it connects the feature mapping from these dual ways in spatial sizes and then variations the sizes to the unique Clr utilizing the 1×1 convolutional transformation. Then, it uses Swish activation and batch normalization processes to get the middle feature mapping comprising information from either direction, as the equation below exposes.

$$f = \delta \left(F_1 \left(\left[\frac{1}{W} \sum_{0 \leq j \leq W} x_c(h, j), \frac{1}{H} \sum_{0 \leq i \leq H} x_c(i, w) \right] \right) \right) \quad (1)$$

Here, f denotes intermediate feature mapping gained by encoder spatial information in dual ways, δ represents the Swish activation function, and F_1 refers to the function of convolution transformation of 1×1 . Now, x_c denotes feature data of the particular location of the feature graph in channel c , h denotes the specific height of feature mapping, and j symbolizes feature mapping width by the value range between $[0, W]$. Also, w represents the particular width of the feature mapping, and i mean feature mapping height through the value ranges from $[0, H]$. F can be separated into dual tensors, h^f and w^f , along with the spatial dimensions in dual ways. Over dual 1×1 convolution transformation functions, h^f and w^f are transformed into tensors by

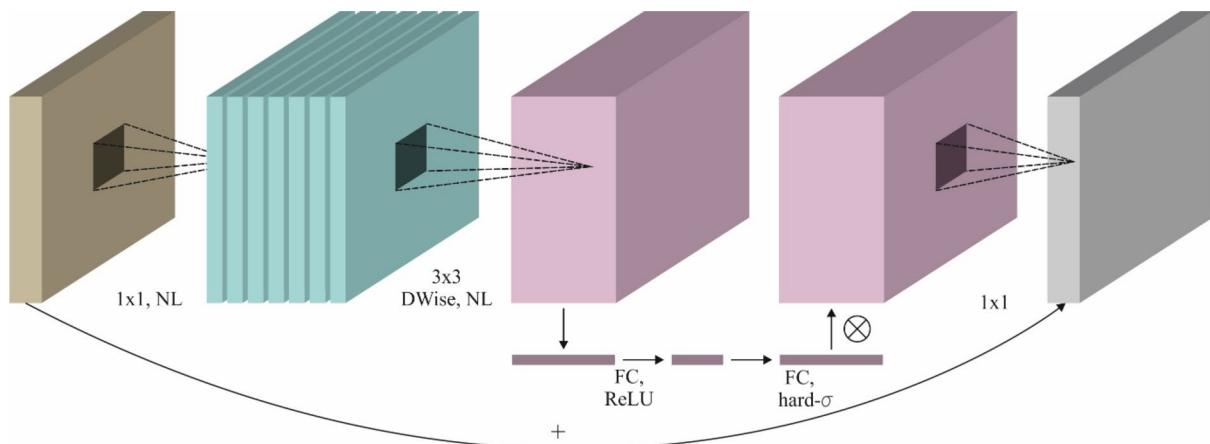


Fig. 3. MobileNetV3 architecture.

the equivalent channel counts as the input X . Lastly, it multiplies the lengthy attention weight with X to obtain the CA module output, as the equation below shows.

$$y_c = x_c(i, j) \cdot (\sigma [F_h(f^h)]) \cdot (0 [F_w(f^w)]) \quad (2)$$

While y_c denotes the output of the c th channel, 0 means the activation function of the sigmoid, and F_h and F_w represent convolution transformation functions in width and height.

Classification process: hybrid DL models

For the SLR process, the hybrid of the CNN-BiGRU-A classifier is employed³¹. This hybrid model was chosen for its ability to handle spatial and temporal information effectively, which is significant for accurate SLR. CNN outperforms in extracting spatial features from images, while BiGRU captures sequential dependencies, making it ideal for comprehending the temporal aspect of sign gestures. Adding AMs allows the model to concentrate on the most crucial features in a sequence, improving recognition accuracy by mitigating noise and irrelevant data. This integration enables the model to process dynamic, real-world SL data more effectually than conventional methods that may only focus on one aspect (spatial or temporal) at a time. Furthermore, this hybrid methodology ensures that the model can handle the complexity and variability of SL gestures, giving superior performance to simpler architectures. The integration of these techniques presents a robust solution to the challenges in SLR, particularly for continuous and dynamic gestures. Figure 4 portrays the structure of the CNN-BiGRU-A model.

The CNN-BiGRU-A method comprises 3 core elements. Initially, CNN is applied to remove local temporal features from the time sequences subsiding information, assisting the process in recognizing short-term forms within the data through various monitoring scores. Bi-GRU handles longer-range dependency in the time sequences, permitting the method to consider previous or upcoming subsiding tendencies, which develops complete prediction precision. Finally, the AM concentrates on the most significant time intervals, allocating high weight to important moments of change and improving the performance of the models by ordering primary data. This mixture allows the method to capture composite subsiding patterns successfully and makes precise predictions. For instance, in a mining region using composite subsiding behaviour, CNN identifies fast, localized variations at different observing points. BiGRU then trajectories longer-range tendencies by combining historical or present data, assisting the model in identifying gradually growing subsiding patterns. The AM emphasizes moments of abrupt change, leading the model's concentration to crucial changes, like abrupt growths in subsiding rate. Mutually, these modules guarantee timely and precise predictions, making the method helpful in dynamical mining atmospheres.

The CNN module contains various layers, which work together to remove essential patterns from the input data. This convolution layer recognizes particular attributes within the data by calculating weighting amounts, whereas pooling and activation layers present nonlinearity and decrease data sizes, effectively allowing the system to identify composite subsiding patterns. Normalization and fully connected (FC) layers enhance the last predictions by normalization, enhancing training speed or model strength. The major equations are as shown:

$$(I * K)_{ij} = \sum_m \sum_n I_{m+i, n+j} \cdot K_{mn} \quad (3)$$

$$P_{ij} = \max(I_{i-m, j-n}) \quad (4)$$

$$O = \sigma(W \cdot I + b) \quad (5)$$

$$\hat{x} = \frac{x - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}}, y = \gamma \hat{x} + \beta \quad (6)$$

During Eq. (3), $(I * K)_{ij}$ signifies the output feature mapping value at location (i, j) after the convolutional process. I denotes input data, K represents the convolutional kernel, and i, j , and m, n represent position indices over output feature mapping and the convolutional kernel. In Eq. (4), P_{ij} characterizes the output feature

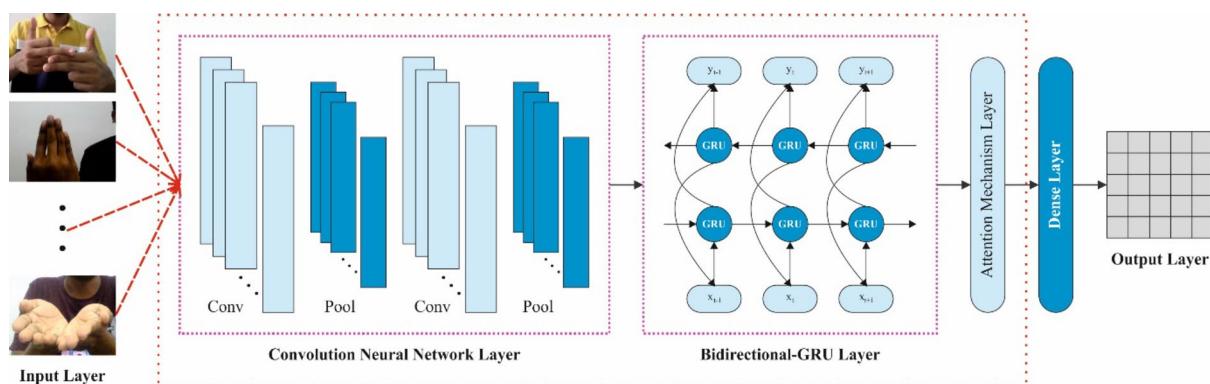


Fig. 4. Structure of CNN-BiGRU-A method.

mapping value at location (i, j) after the pooling process. Also, m and n represent position indices of the pooling window. During Eq. (5), O refers to output, I means input features, W stands for weighted matrix, b signifies biased matrix, and σ denotes activation function. Equation (6) characterizes the normalization layer task, while x signifies input data, \hat{x} symbolizes the standardized input data, σ_B^2 and μ_B means the variance and mean of the present minibatch data, correspondingly. ϵ denotes constant for numerical accuracy, whereas γ and β are learnable parameters.

The GRU module enhances prediction precision by controlling the flow of information over its update and reset gates, which define the related data to keep or discard at all steps. This mechanism permits the method to effectively take important patterns in subsiding information regarding either recent inputs or previous observations. For instance, in predicting subsiding tendencies, the GRU utilizes previous and present data scores to recognize a steady pattern that improves the model's capability for predicting upcoming subsiding precisely. Applying the hidden layer (HL) h_{t-1} from the preceding time step and the present input x_t , the GRU approach can be signified as shown:

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r) \quad (7)$$

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t] + b_z) \quad (8)$$

Here, W_r , W_z characterizes the weighted matrices, b_r , b_z symbolize the biased vectors, and σ means the sigmoid activation function. The reset gate r_t defines which data from the preceding HL h_{t-1} must be discarded; however, the update gate z_t selects the mixing ratio of the newer and older memories.

Then, the last output is gained by computing the candidate HL \tilde{h}_t and the HL h_t . The HL is then distributed to another layer or applied as the previous output.

$$\bar{h} = \tanh(W \cdot [r_t \odot h_{t-1}, x_t] + b) \quad (9)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (10)$$

Here W denotes the weighted matrix, \tilde{h}_t can be approximated utilizing x_t and r_t to gain the promising HL. Lastly, \tilde{h}_t and h_t are weighted to obtain the last condition fusion degree, \odot recognizes the Hadamard functions.

This bidirectional model allows the method to recognize patterns within the data more efficiently, which is helpful in tracking modifications that subside over time.

The AM enhances this ability by specially targeting significant portions of the data. It allocates greater weight to crucial data at every step, selecting main characteristics that might specify essential variations. For instance, in a real-time situation, Bi-GRU's main characteristics are tendencies in previous or upcoming contexts, whereas the AM highlights unexpected moves or crucial points within the information, like regions where rates quickly rise. This integration permits the method to better adjust to useful requirements in mining regions using wide-ranging behaviours. These particular equations are as shown in (11)-(13) :

$$\alpha_{ij} = \frac{\exp(score(h_i, \bar{h}_j))}{\sum \exp(score(h_n, \bar{h}_m))} \quad (11)$$

$$c_i = \sum \alpha_{ij} \bar{h}_S \quad (12)$$

$$\alpha_i = f(c_i, h_i) = \tanh(W_c \cdot [c_i, h_i]) \quad (13)$$

Now, α_{ij} characterizes the attention score computed amongst the output of the encoder at the j^{th} time step and the decoding layer at i^{th} time steps. h signifies the HL at all-time steps, W symbolizes the weighted matrix related to the input or HL, and α_i signifies the final attention weighting gained over the AM.

This prediction process includes three significant stages: Initially, data can be pre-processed in the CNN layer over pooling and convolution to make feature-rich data. Next, these vectors are given to a layer of BiGRU that takes either short- or long-term designs within the data and avoids gradient problems. Lastly, the AM allocates weight to the main features, decreasing unrelated data and enhancing model efficacy. This allows the method to concentrate on essential patterns in subsiding data, leading to precise predictions.

Parameter optimizing process: AROA

Finally, the AROA optimally adjusts the hyperparameter values of the CNN-BiGRU-A approach, resulting in more excellent classification performance³². This method was chosen for parameter optimization due to its ability to balance exploration and exploitation during the optimization process effectively. Unlike conventional optimization methods that may get stuck in local minima, AROA uses attraction and repulsion mechanisms to explore the solution space more thoroughly and avert suboptimal solutions. This is beneficial for DL models with complex parameter spaces. Furthermore, AROA's simplicity and efficiency make it a robust choice for optimizing resource-intensive models like DL networks without needing extensive computational power. The algorithm's flexibility in fine-tuning hyperparameters enhances model convergence and accuracy, improving overall performance. The ARO model's adaptability and capability to optimize parameters such as learning rates, batch sizes, and network architecture make it superior to more conventional optimization techniques like grid or

random search. AROA is a practical and effectual solution for improving the model's performance in SLR tasks. Figure 5 demonstrates the structure of the AROA model.

This method naturally imitates the phenomenon of attraction-repulsion. The initial phase in AROA is to make the first value of n individuals X .

$$X_i = \text{rand} \odot (X_{up} - X_{low}) + X_{low} \quad (14)$$

During Eq. (14), X_i refers to the value of the i^{th} individual, X_{low} and X_{up} mean lower and upper limits of the searching space, individually. rand denotes the randomly generated vector.

Then, each fitness value X_i is calculated and the best is defined based on the testing problem. The following stage in AROA utilizes the theory of attraction and repulsion, which relies on the distance between individuals X . Hence, the value of X can be upgraded by computing the fitness levels of neighbouring individuals. The distance between i^{th} and j^{th} has been calculated as shown:

$$D = \begin{bmatrix} d_{1,1} & d_{1,2} & d_{1,3} & \dots & d_{1,n} \\ d_{2,1} & d_{2,2} & d_{2,3} & \dots & d_{2,n} \\ d_{3,1} & d_{3,2} & d_{3,3} & \dots & d_{3,n} \\ \dots & \dots & \dots & \dots & \dots \\ d_{n,1} & d_{n,2} & d_{n,3} & ? & d_{n,n} \end{bmatrix} \quad (15)$$

$$d^2(X_i, X_j) = \sum_{k=1}^{\text{dim}} (x_i^k - x_j^k)^2 \quad k = 1 \text{ to } \text{dim} \quad (16)$$

Meanwhile, X_i and X_j correspondingly provide the values of i^{th} and j^{th} individuals, and dim denotes X_i 's dimension counts.

The following operation is to update the attraction-repulsion operator (n_i) according to the distance from the i^{th} individual to the furthermost member of X ($d_{i,\max}$) and $d_{i,j} \in D$. This can be described as shown:



Fig. 5. Structure of AROA method.

$$n_i = \frac{1}{n \sum_{j=1}^k (X_j - X_i)} \cdot \left(1 - \frac{d_{i,j}}{d_{i,\max}}\right) \cdot s(f_i, f_j). \quad (17)$$

Here, c stands for the step size, and s signifies the function, which controls the direction of the change based on fitness value, and the s value is upgraded as:

$$s(f_i, f_j) = \begin{cases} 1 & f_i > f_j \\ 0 & f_i = f_j \\ -1 & f_i < f_j \end{cases} \quad (18)$$

Additionally, the k value, in Eq. (17), denotes the neighbour number, which reduces with excess the iterations and it can be upgraded as:

$$k = ? \left(1 - \frac{t}{t_{\max}}\right) \cdot n? + 1 \quad (19)$$

Now, t means present iteration and t_{\max} stands for maximal iteration counts.

The next step is to utilize attraction to determine the optimal solution. This process characterizes the exploration stage as equivalent to other MH models determining the possible area. The Attraction operator () can be described as:

$$b_i = \begin{cases} c \cdot m \cdot (X_{best} - X_i) r_1 \geq p_1 \\ c \cdot m \cdot (a_1 X_{best} - X_i) r_1 < p_1 \end{cases} \quad (20)$$

Whereas X_{best} signifies optimal solution, a_1 designates randomly generated vectors. The parameter $r_1 \in [0,1]$ means a randomly generated number, and p_1 indicates probability thresholds. The parameter m has been utilized to mimic the impact of the best solution, and it is necessary for controlling the balance between exploitation and exploration; it is outlined as follows:

$$m = \frac{1}{2} \left(\frac{\exp\left(18 \cdot \left(\frac{t}{t_{\max}}\right) - 4\right) - 1}{\exp\left(18 \cdot \left(\frac{t}{t_{\max}}\right) - 4\right) + 1} + 1 \right) \quad (21)$$

Consequently, the exploration phase of AROA has been employed to improve the probability of defining the optimal solution. This process can be described as shown:

$$X_i(t) = X_i(t-1) + n_i + b_i + r_i, \quad (22)$$

$$r_i = \begin{cases} r_B & r_3 > 0.5 \cdot \frac{t}{t_{\max}} + 0.25 \\ r_{tri} & r_3 \leq 0.5 \cdot \frac{t}{t_{\max}} + 0.25 \\ r_R & r_2 < p_2 \\ & r_2 \geq p_2 \end{cases} \quad (23)$$

While r_B represents the operator that signifies the Brownian motion by upgrading the standard deviation based on the searching area limits, and it can be described as:

$$r_B = u_1 \odot N\left(0, f r_1 1 - \frac{t}{t_{\max}} \cdot (X_{up} - X_{low})\right) \quad (24)$$

Whereas u_1 denotes a binary vector. N signifies the randomly generated vector value after a normal distribution, and $f r_1$ symbolizes the contact value.

Besides, r_{tri} denotes the second operator, which relies on trigonometric functions and the individual that can be chosen using the roulette wheel selection. This can be outlined as shown:

$$r_{tri} = \begin{cases} fr_2 \cdot u_2 \cdot \left(1 - \frac{t}{t_{\max}}\right) \cdot \sin(2r_5\pi) \odot |a_2 \odot X_w - X_i| & r_4 < 0.5 \\ fr_2 \cdot u_2 \cdot \left(1 - \frac{t}{t_{\max}}\right) \cdot \cos \odot (2r_5\pi) |a_2 \odot X_w - X_i| & r_4 \geq 0.5 \end{cases} \quad (25)$$

Here, fr_2 denotes the multiplier, u_2 refers to binary vectors, and r_4 and r_5 are randomly generated numbers amongst (0, 1). a_2 relates to a randomly generated vector comprising values ranging between (0-1). X_w denotes a randomly chosen solution from X .

During Eq. (23), r_R refers to the third operator applied to improve the value of X_i , and it can be well-defined as:

$$r_R = u_3 \odot (2 \cdot a_3 - o) \odot (X_{pll} - X_{low}) \quad (26)$$

u_3 denotes the binary vector gained using the threshold tr_3 used for every solution. a_3 indicates randomly selected vector values, and zero stands for matrix unit.

Additionally, the eddy formation theory can be used to improve the solution, and this can be expressed as:

$$X_i = \begin{cases} X_i + c_{fp} (u_4 (a_4 (X_{ll} - X_{low}) + X_{low})) & r_6 < e_f \\ X_i + (e_f \cdot (1 - r_7) + r_7) (X_{r8} - X_{r9}) & r_6 \geq e_f \end{cases} \quad (27)$$

Here, r_7 signifies randomly formed integers variant from (0-1), and e_f signifies probability cutoffs. u_4 indicates a binary vector gained by the threshold of $1 - e_f$, and a_4 represents a vector containing arbitrary numbers. However, r_8 and r_9 are agent indexes randomly selected from X , and c_f means parameter upgraded as shown:

$$c_f = \left(1 - \frac{t}{t_{\max}}\right)^3 \quad (28)$$

After, the memory is measured as the subsequent impact applied to upgrade the solutions. This can be directed by comparing the novel value of the solution using its old value and preservative the best of them as expressed in the subsequent Eq. (29).

$$X_i(t) = \begin{cases} X_i(t) & f(X_i(t)) < f(X_i(t-1)) \\ X_i(t-1) & f(X_i(t)) \geq f(X_i(t-1)) \end{cases} \quad (29)$$

The AROA originates a fitness function (FF) to improve classifier performance. It defines an optimistic number to characterize the higher efficiency of the candidate solution. Here, the decrease in classification rate of error is deliberated as FF. Its formulation is mathematically expressed in Eq. (30).

$$\begin{aligned} \text{fitness}(x_i) &= \text{ClassifierErrorRate}(x_i) \\ &= \frac{\text{no. of misclassified samples}}{\text{Total no. of samples}} \times 100 \end{aligned} \quad (30)$$

Performance validation

The SACHI-SLRHDL model is examined under an ISL dataset³³. This dataset consists of 20 class labels below 800 images, as shown in Table 1. Figure 6 represents the sample images.

Evaluation metrics for classification models: Acc_y , Prec_n , Recal_l , F_{Score} , and MCC
The performance of the classification models is evaluated using diverse metrics. Equation (31) represents acc_y , which measures the overall proportion of correct predictions. Equation (32) computes prec_n , the ratio of correct positive predictions. Equation (33) defines recal_l , which evaluates the capability of the model to detect true positives. Equation (34) shows the F_{Score} , a metric that integrates prec_n and recal_l into a single value to balance their trade-offs, particularly in cases with imbalanced classes. Finally, Eq. (35) represents MCC , which evaluates the balance between classification accuracy for both classes, providing a more balanced performance

Phrases	Class labels	Images for phrases
Again	C-1	40
Agree	C-2	40
Answer	C-3	40
Book	C-4	40
Break	C-5	40
Careful	C-6	40
chat	C-7	40
Email	C-8	40
Good morning	C-9	40
Happy birthday	C-10	40
How are you	C-11	40
Hungry	C-12	40
I Need Help	C-13	40
Practice	C-14	40
Problem	C-15	40
Question	C-16	40
Sick	C-17	40
Sun	C-18	40
Together	C-19	40
Wait	C-20	40
Total number of images		800

Table 1. Details of dataset.



Fig. 6. Sample images.

evaluation in imbalanced datasets. These metrics give a comprehensive assessment of model performance as represented by the following equations:

$$Accu_y = \frac{TP + TN}{TP + TN + FP + FN} \quad (31)$$

$$Prec_n = \frac{TP}{TP + FP} \quad (32)$$

$$Recal_l = \frac{TP}{TP + FN} \quad (33)$$

$$F_{Score} = \frac{2 \cdot Prec_n \cdot Recal_l}{Prec_n + Recal_l} \quad (34)$$

$$MCC = \frac{(TP.TN) - (FP.FN)}{\sqrt{(TP + FP) . (TP + FN) . (TN + FP) . (TN + FN)}} \quad (35)$$

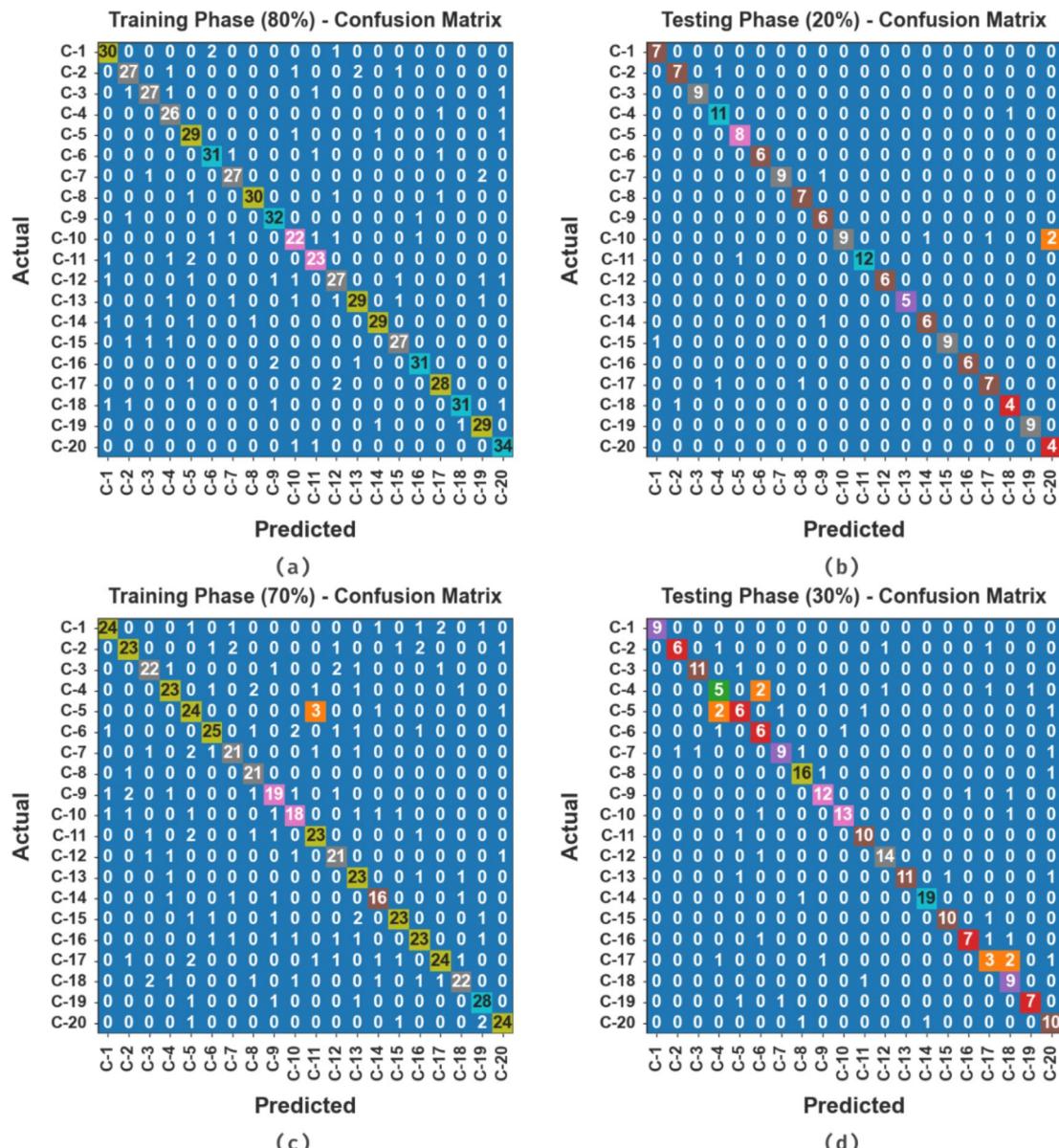
Here, TP represents True Positives, TN indicates True Negatives, FP denotes False Positives, and FN stands for False Negatives. These metrics provide a comprehensive evaluation of the model's performance, including its capability to correctly detect both positive and negative instances and its handling of class imbalances. The inclusion of $prec_n$, $recal$, F_{Score} , and MCC presents insights into the efficiency of the model in detecting relevant patterns while minimizing errors across diverse classification scenarios.

Analysis and evaluation of experimental results

Figure 7 states the confusion matrix generated through the SACHI-SLRHDL approach below 80%:20% and 70%:30% of TRAPS/TEPS. The performances designate that the SACHI-SLRHDL model has effectual detection and identification of 20 classes accurately.

Table 2; Fig. 8 illustrate the SL detection of the SACHI-SLRHDL approach below 80%TRAPS and 20%TEPS. The table values implied that the SACHI-SLRHDL approach has attained efficient performance. Using 80%TRAPS, the SACHI-SLRHDL approach gains typical $accu_y$ of 98.89%, $prec_n$ of 88.89%, $recal$ of 88.85%, F_{Score} of 88.80%, and MCC of 88.25%, correspondingly. Moreover, using 20%TRAPS, the SACHI-SLRHDL method obtains standard $accu_y$ of 99.19%, $prec_n$ of 91.54%, $recal$ of 93.21%, F_{Score} of 91.87%, and MCC of 91.72%, correspondingly.

Table 3; Fig. 9 demonstrate the SL detection of the SACHI-SLRHDL approach below 70%TRAPS and 30%TEPS. The values of the table implied that the SACHI-SLRHDL approach has gained efficient performance. Using 70%TRAPS, the SACHI-SLRHDL method obtains typical $accu_y$ of 97.98%, $prec_n$ of 79.93%, $recal$ of



Class labels	<i>Accu_y</i>	<i>Prec_n</i>	<i>Recal_l</i>	<i>F_{score}</i>	<i>MCC</i>
TRAPS (80%)					
C-1	98.91	88.24	90.91	89.55	88.99
C-2	98.59	87.10	84.38	85.71	84.99
C-3	98.91	90.00	87.10	88.52	87.96
C-4	98.91	83.87	92.86	88.14	87.69
C-5	98.59	82.86	90.62	86.57	85.92
C-6	99.06	91.18	91.18	91.18	90.68
C-7	99.06	90.00	90.00	90.00	89.51
C-8	99.38	96.77	90.91	93.75	93.47
C-9	99.06	88.89	94.12	91.43	90.97
C-10	98.44	81.48	81.48	81.48	80.67
C-11	98.75	85.19	85.19	85.19	84.53
C-12	97.97	81.82	79.41	80.60	79.54
C-13	98.59	90.62	82.86	86.57	85.92
C-14	99.06	93.55	87.88	90.62	90.18
C-15	99.06	90.00	90.00	90.00	89.51
C-16	99.22	93.94	91.18	92.54	92.14
C-17	99.06	90.32	90.32	90.32	89.83
C-18	99.22	96.88	88.57	92.54	92.23
C-19	99.06	87.88	93.55	90.62	90.18
C-20	98.91	87.18	94.44	90.67	90.17
Average	98.89	88.89	88.85	88.80	88.25
TESPS (20%)					
C-1	99.38	87.50	100.00	93.33	93.24
C-2	98.75	87.50	87.50	87.50	86.84
C-3	100.00	100.00	100.00	100.00	100.00
C-4	98.12	84.62	91.67	88.00	87.07
C-5	99.38	88.89	100.00	94.12	93.97
C-6	100.00	100.00	100.00	100.00	100.00
C-7	99.38	100.00	90.00	94.74	94.55
C-8	99.38	87.50	100.00	93.33	93.24
C-9	98.75	75.00	100.00	85.71	86.04
C-10	97.50	100.00	69.23	81.82	82.10
C-11	99.38	100.00	92.31	96.00	95.75
C-12	100.00	100.00	100.00	100.00	100.00
C-13	100.00	100.00	100.00	100.00	100.00
C-14	98.75	85.71	85.71	85.71	85.06
C-15	99.38	100.00	90.00	94.74	94.55
C-16	100.00	100.00	100.00	100.00	100.00
C-17	98.12	87.50	77.78	82.35	81.52
C-18	98.75	80.00	80.00	80.00	79.35
C-19	100.00	100.00	100.00	100.00	100.00
C-20	98.75	66.67	100.00	80.00	81.12
Average	99.19	91.54	93.21	91.87	91.72

Table 2. SL detection of SACHI-SLRHDL approach under 80%TRAPS and 20%TESPS.

80.04%, F_{score} of 79.76%, and MCC of 78.83%, respectively. Additionally, using 30%TRAPS, the SACHI-SLRHDL technique reaches a typical $accu_y$ of 98.04%, $prec_n$ of 79.32%, $recal_l$ of 78.56%, F_{score} of 78.44%, and MCC of 77.69%, respectively.

Figure 10 depicts the training (TRA) $accu_y$ and validation (VAL) $accu_y$ performances of SACHI-SLRHDL technique below 80%TRAPS and 20%TESPS. The $accu_y$ values are computed across an interval of 0–50 epochs. The figure identified that the values of TRA and VAL $accu_y$ show an increasing trend, indicating the capability of the SACHI-SLRHDL approach through maximum performance across multiple repetitions. Furthermore, the TRA and VAL $accu_y$ values remain close through the epochs, notifying lesser overfitting and showcasing the higher performance of the SACHI-SLRHDL approach, which guarantees reliable prediction over hidden samples.

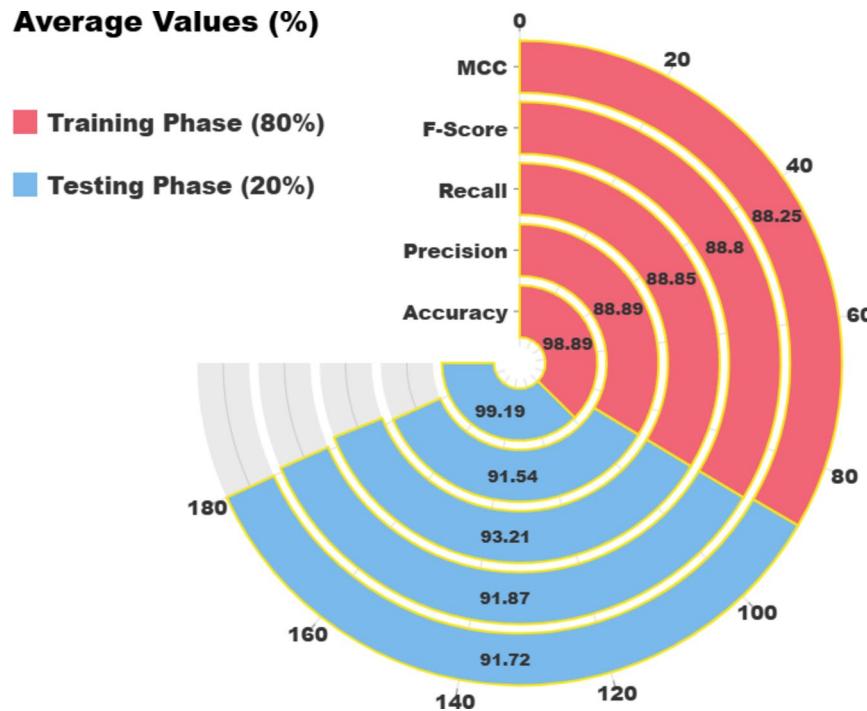


Fig. 8. Average of SACHI-SLRHDL approach under 80%TRAPS and 20%TESPS.

Figure 11 shows the TRA loss (TRALOS) and VAL loss (VALLOS) graph of the SACHI-SLRHDL approach with 80%TRAPS and 20%TESPS. The loss values are computed through an interval of 0–50 epochs. The values of TRALOS and VALLOS demonstrate a reducing trend, which indicates the proficiency of the SACHI-SLRHDL approach in harmonizing an exchange between generalization and data fitting. The continual reduction in loss values assures the superior performance of the SACHI-SLRHDL method and tuning of the prediction results afterwards.

In Fig. 12, the PR curve inspection of the SACHI-SLRHDL approach below 80%TRAPS and 20%TESPS offers an understanding of its outcome by scheming Precision instead of Recall for 20 distinct class labels. The figure exhibits that the SACHI-SLRHDL technique continually attains enhanced PR values over distinct class labels, which indicates its proficiency in keeping a high proportion of true positive predictions (precision) while effectively grabbing a significant share of actual positives (recall).

Figure 13 examines the ROC outcome of the SACHI-SLRHDL approach below 80%TRAPS and 20%TESPS. The performances showed that the SACHI-SLRHDL approach gains superior ROC analysis across each class label, representing noteworthy proficiency in understanding the classes. This consistent tendency of maximum ROC values through several classes illustrates the skilful outcomes of the SACHI-SLRHDL technique in predictive classes, implying the classification system's robust nature.

Comparative analysis of the SACHI-SLRHDL model performance across different techniques under the ISL dataset

Table 4; Fig. 14 compare the SACHI-SLRHDL method's comparison performances with the existing techniques^{19,34,35}. The performances highlight that the DMM-MobileNet, Bi-SRN, Skeletal Feature + LSTM, ANFIS Networks, MLP-MDC, PCNN, Modified K-NN, Bi-LSTM, TCNN, CNN-BiLSTM, and DCNN methodologies have exhibited poorer performance. Likewise, SVM models have attained closer outcomes by $prec_n$, $recal$, $accu_y$, and F_{score} of 89.04%, 89.53%, 98.26%, and 89.30%, respectively. Additionally, the SACHI-SLRHDL technique reported enhanced performance with higher $prec_n$, $recal$, $accu_y$, and F_{score} of 91.54%, 93.21%, 99.19%, and 91.87%, respectively.

Comparative evaluation of computational time for the SACHI-SLRHDL model across different techniques under the ISL dataset

Table 5; Fig. 15 depict the computational time (CT) analysis of the SACHI-SLRHDL technique compared to existing methods. The SACHI-SLRHDL model demonstrates the fastest CT at 6.98 s. Compared to other methods like DMM-MobileNet at 22.61 s and Bi-SRN method at 22.37 s, the SACHI-SLRHDL model significantly outperforms, suggesting its effectualness in real-time applications. While models such as Bi-LSTM at 17.15 s and Skeletal Feature plus LSTM at 13.74 s are faster than some conventional methods, the SACHI-SLRHDL method stands out in terms of minimizing CT without sacrificing performance. Other models, comprising PCNN method at 24.67 s and CNN-BiLSTM at 23.24 s, require more time, highlighting the SACHI-SLRHDL model's advantage in speed. The efficiency of the SACHI-SLRHDL model makes it specifically appropriate for applications with low-latency requirements, presenting an optimal balance between performance and computational cost.

Class labels	<i>Accu_y</i>	<i>Prec_n</i>	<i>Recal_i</i>	<i>F_{score}</i>	<i>MCC</i>
TRAPS (70%)					
C-1	98.21	88.89	77.42	82.76	82.04
C-2	97.68	82.14	74.19	77.97	76.85
C-3	98.04	81.48	78.57	80.00	78.98
C-4	97.86	79.31	79.31	79.31	78.18
C-5	97.14	68.57	82.76	75.00	73.86
C-6	97.86	83.33	78.12	80.65	79.56
C-7	98.04	80.77	77.78	79.25	78.23
C-8	98.75	77.78	95.45	85.71	85.56
C-9	97.50	73.08	73.08	73.08	71.77
C-10	97.68	75.00	72.00	73.47	72.27
C-11	97.86	79.31	79.31	79.31	78.18
C-12	98.04	75.00	84.00	79.25	78.36
C-13	97.86	71.88	88.46	79.31	78.66
C-14	98.39	76.19	80.00	78.05	77.24
C-15	98.21	85.19	79.31	82.14	81.26
C-16	97.50	76.67	76.67	76.67	75.35
C-17	97.86	85.71	75.00	80.00	79.07
C-18	97.86	84.62	73.33	78.57	77.67
C-19	98.57	84.85	90.32	87.50	86.79
C-20	98.75	88.89	85.71	87.27	86.63
Average	97.98	79.93	80.04	79.76	78.83
TESPS (30%)					
C-1	100.00	100.00	100.00	100.00	100.00
C-2	98.33	85.71	66.67	75.00	74.78
C-3	99.17	91.67	91.67	91.67	91.23
C-4	95.42	50.00	45.45	47.62	45.28
C-5	96.25	60.00	54.55	57.14	55.26
C-6	97.08	54.55	75.00	63.16	62.53
C-7	97.50	81.82	69.23	75.00	73.98
C-8	97.92	84.21	88.89	86.49	85.40
C-9	97.92	80.00	85.71	82.76	81.71
C-10	98.75	92.86	86.67	89.66	89.05
C-11	98.75	83.33	90.91	86.96	86.39
C-12	98.75	87.50	93.33	90.32	89.71
C-13	98.75	100.00	78.57	88.00	88.06
C-14	99.17	95.00	95.00	95.00	94.55
C-15	99.17	90.91	90.91	90.91	90.47
C-16	98.33	87.50	70.00	77.78	77.44
C-17	96.25	42.86	37.50	40.00	38.16
C-18	97.50	64.29	90.00	75.00	74.88
C-19	98.75	87.50	77.78	82.35	81.86
C-20	97.08	66.67	83.33	74.07	73.06
Average	98.04	79.32	78.56	78.44	77.69

Table 3. SL detection of SACHI-SLRHDL approach under 70%TRAPS and 30%TESPS.

Conclusion

In this study, a SACHI-SLRHDL methodology in IoT was presented. The model involved four distinct processes: image pre-processing, improved MobileNetV3 for feature extractor, hybrid DL classification process, and AROA-based parameter tuning. At the primary stage, the SACHI-SLRHDL model utilized BF for image pre-processing to enhance the excellence of the captured images by reducing noise while preserving edges. Next, the improved MobileNetV3 model extracted relevant input image features. For the SLR process, the hybrid of the CNN-BiGRU-A classifier was employed. Finally, the AROA optimally adjusts the CNN-BiGRU-A model's hyperparameter values, resulting in better classification performance. A comprehensive experimental analysis is performed under an ISL dataset to exhibit the more significant solution of the SACHI-SLRHDL method. The experimental validation of the SACHI-SLRHDL method portrayed a superior accuracy value of 99.19% over existing techniques. The limitations of the SACHI-SLRHDL method comprise the restricted size and diversity of

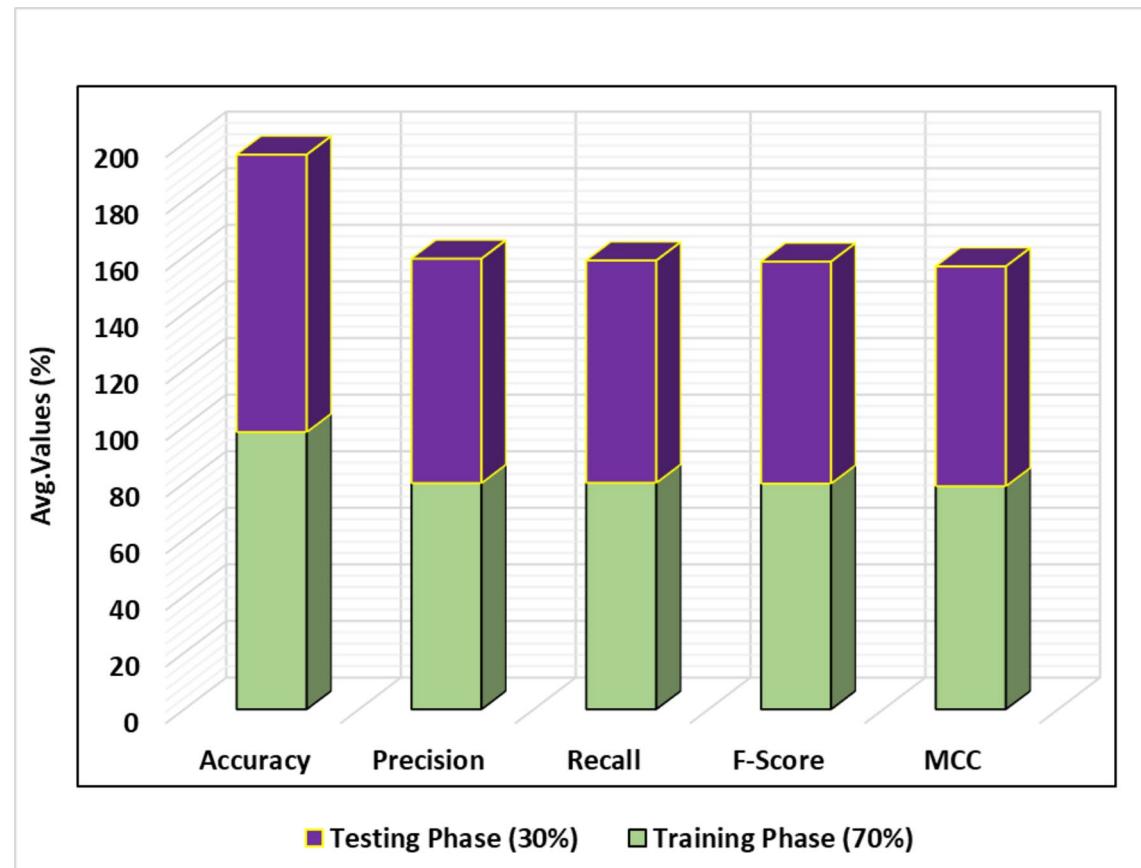


Fig. 9. Average of SACHI-SLRHDL approach under 70%TRAPS and 30%TESPS.

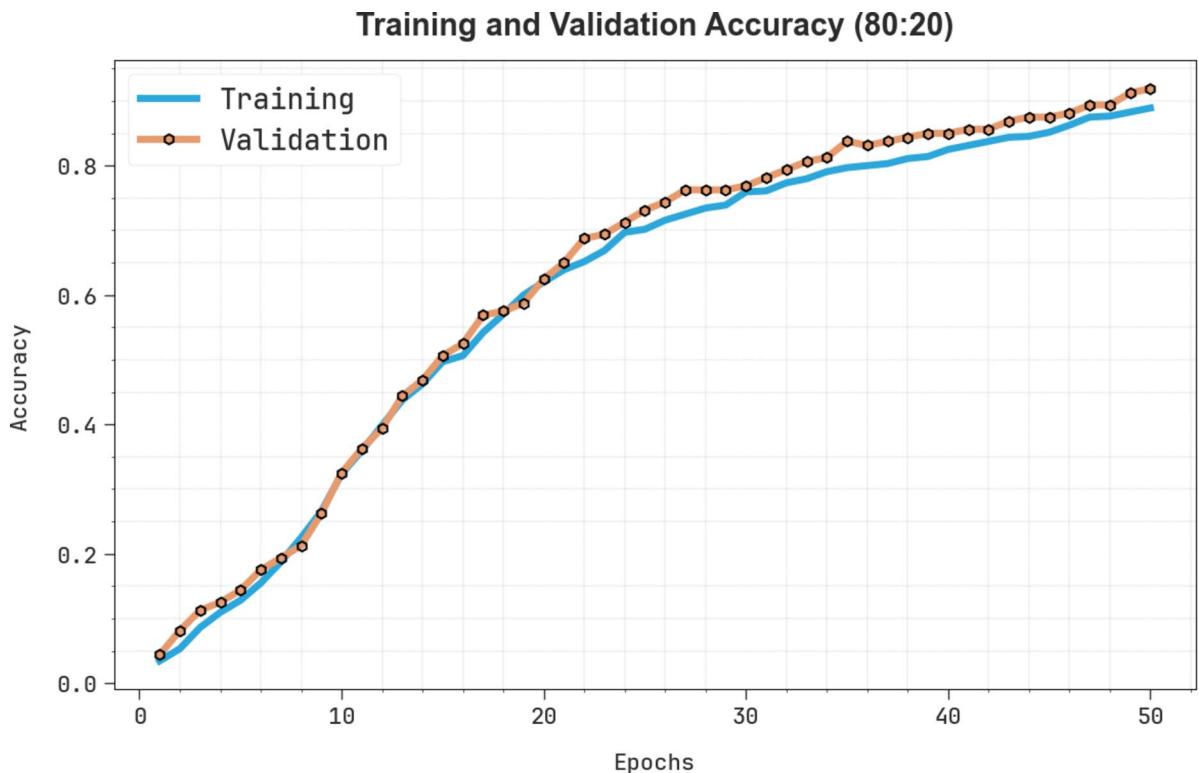
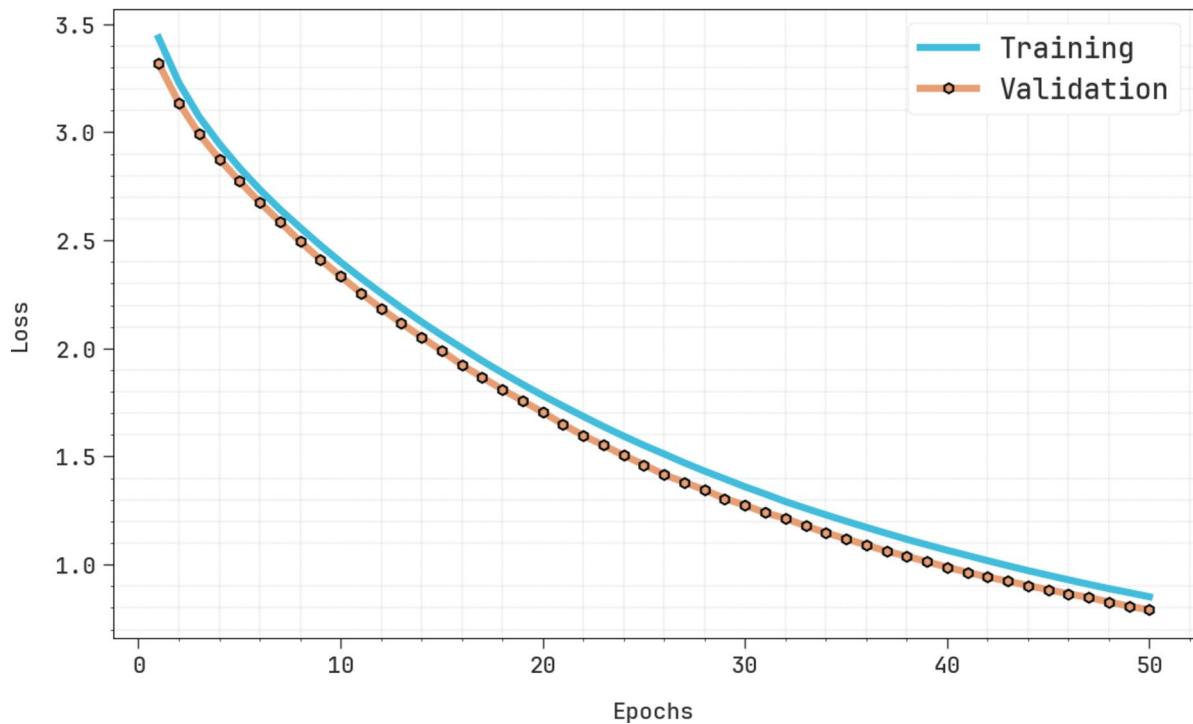
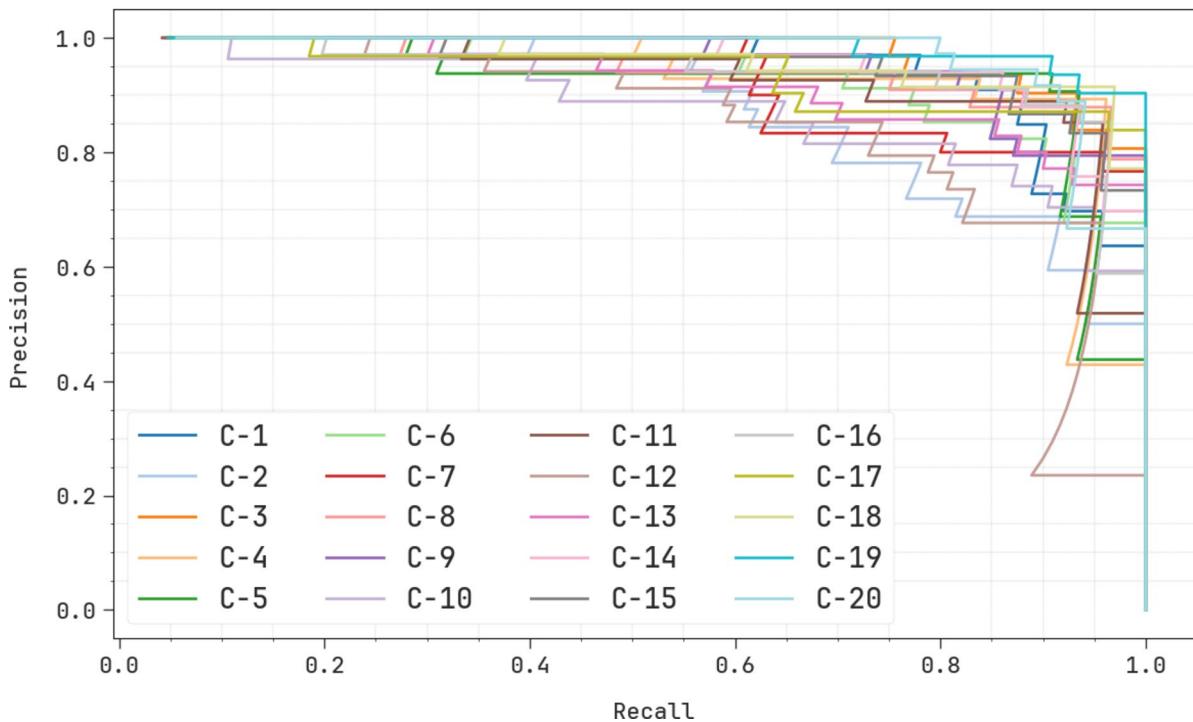


Fig. 10. $Accu_y$ curve of SACHI-SLRHDL approach under 80%TRAPS and 20%TESPS.

Training and Validation Loss (80:20)**Fig. 11.** Loss curve of SACHI-SLRHDL approach under 80%TRAPS and 20%TESPS.**Precision-Recall Curve (80:20)****Fig. 12.** PR curve of SACHI-SLRHDL approach at 80%TRAPS and 20%TESPS.

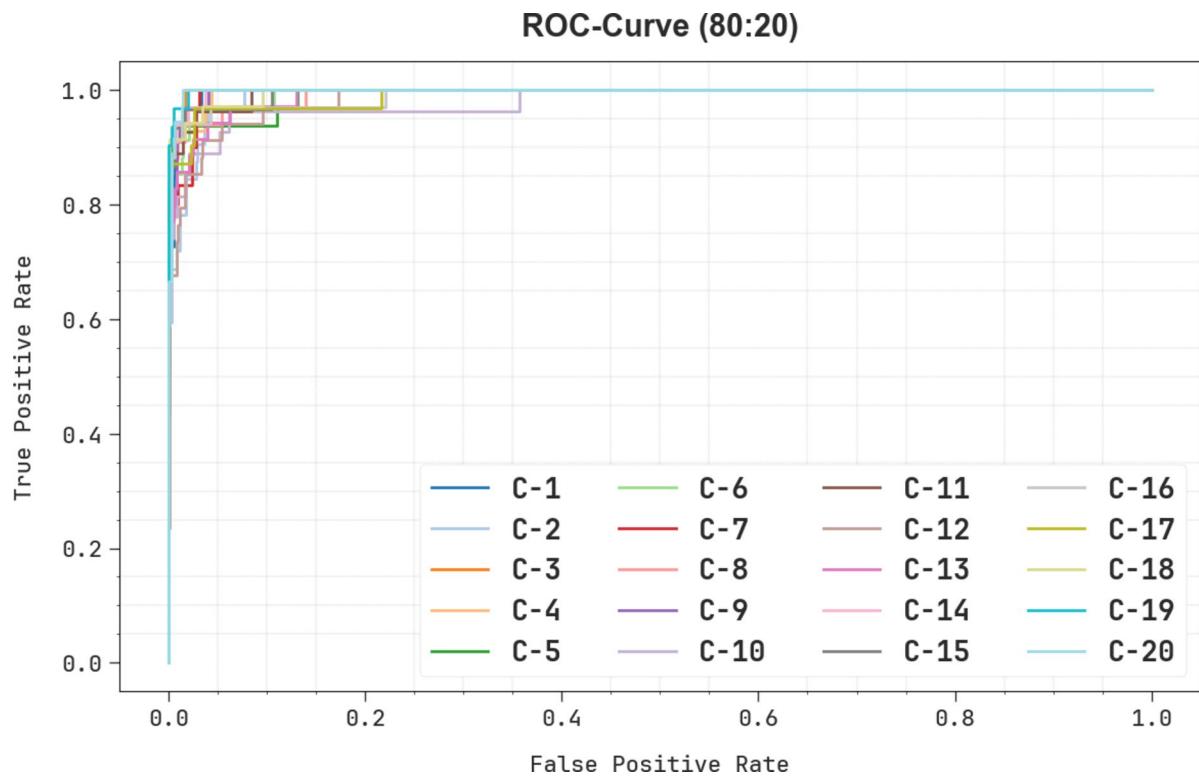


Fig. 13. ROC curve of SACHI-SLRHDL approach under 80%TRAPS and 20%TESPS.

Approach	<i>Accu_y</i>	<i>Precn</i>	<i>Recal</i>	<i>F_{score}</i>
DMM-MobileNet	94.17	89.45	92.25	90.83
Bi-SRN method	97.23	89.09	89.18	89.08
Skeletal feature + LSTM	96.07	89.93	91.03	90.69
ANFIS networks	95.34	89.34	90.10	90.92
MLP-MDC model	98.21	90.29	89.04	90.56
PCNN algorithm	96.27	89.70	91.78	89.49
Modified K-NN	97.32	90.71	91.49	90.85
SVM classifier	98.26	89.04	89.53	89.30
Bi-LSTM	98.29	89.33	90.18	90.07
TCNN	94.39	90.46	90.61	87.85
CNN-BiLSTM	96.09	90.39	89.16	87.68
DCNN	96.77	87.10	88.63	87.00
SACHI-SLRHDL	99.19	91.54	93.21	91.87

Table 4. Comparative outcomes of the SACHI-SLRHDL technique with recent models^{19,34,35}.

the dataset, which may affect the generalization of the model to diverse sign languages and real-world scenarios. Additionally, the performance of the model in challenging environments, such as low lighting or occlusions, has not been fully tested. The computational complexity of the model may hinder its deployment on low-resource devices or IoT platforms. Furthermore, the scalability of the technique for large-scale applications in diverse settings remains unexamined. Future work should concentrate on expanding the dataset to encompass a wider variety of signs and gestures, testing the model under various environmental conditions, and optimizing the approach for deployment on resource-constrained devices. Moreover, further exploration of cross-lingual and cross-cultural adaptability could improve the effectiveness of the model in global applications.

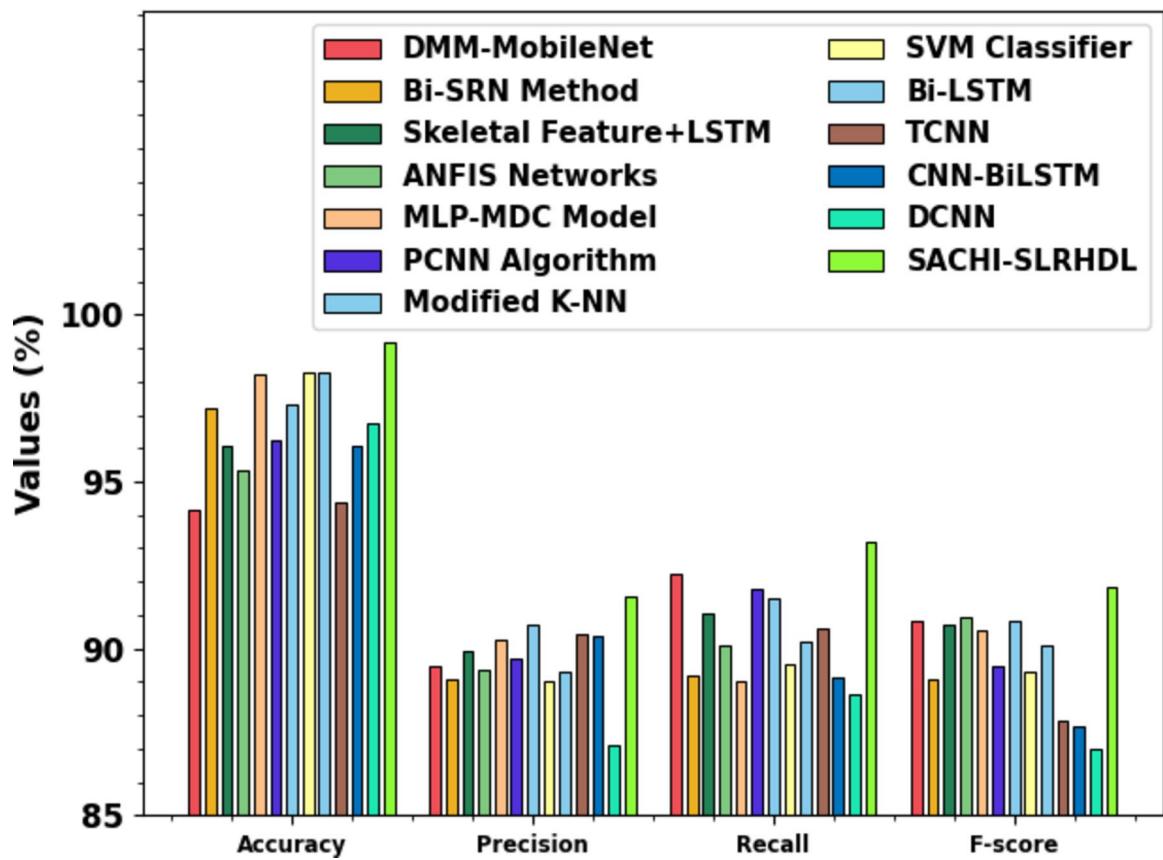


Fig. 14. Comparative outcome of SACHI-SLRHDL technique with recent models.

Approach	CT (sec)
DMM-MobileNet	22.61
Bi-SRN method	22.37
Skeletal feature + LSTM	13.74
ANFIS networks	16.41
MLP-MDC model	18.35
PCNN algorithm	24.67
Modified K-NN	15.24
SVM classifier	14.90
Bi-LSTM	17.15
TCNN	19.08
CNN-BiLSTM	23.24
DCNN	20.09
SACHI-SLRHDL	6.98

Table 5. CT evaluation of the SACHI-SLRHDL technique with existing methods.

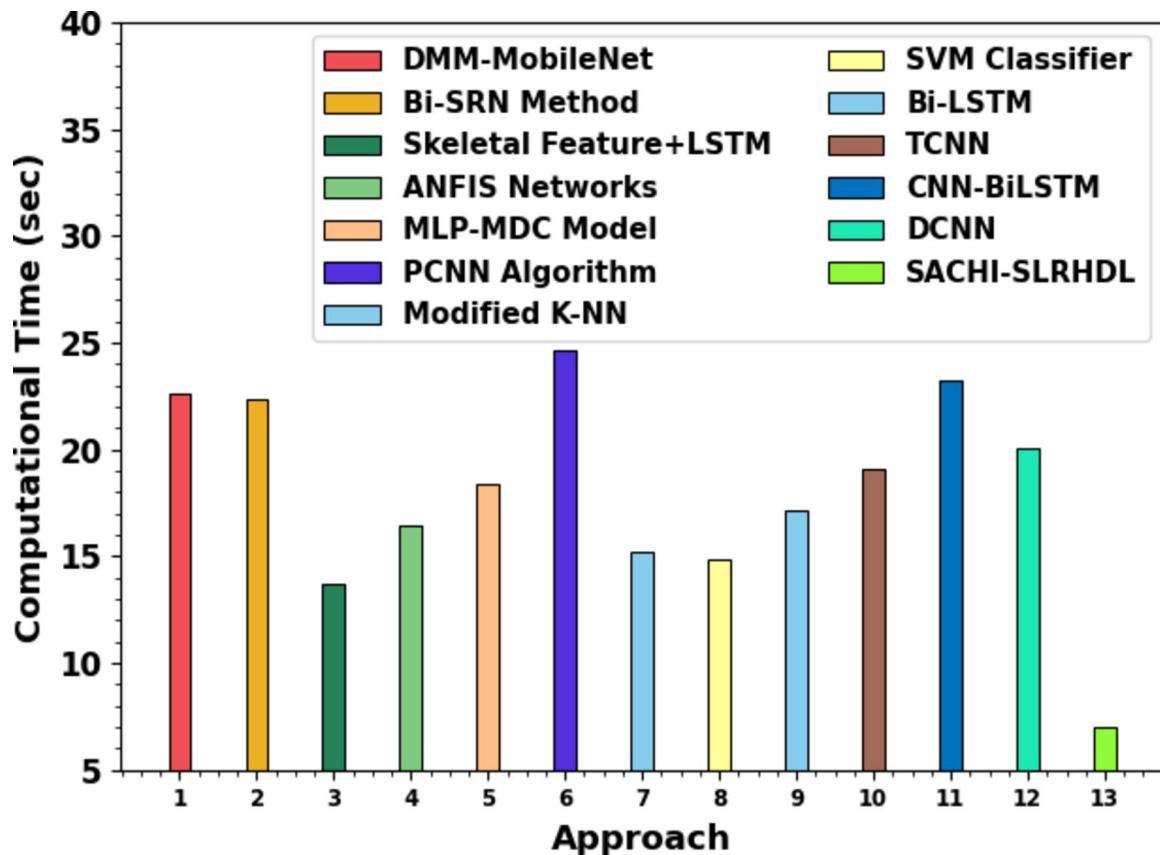


Fig. 15. CT evaluation of the SACHI-SLRHDL technique with existing methods.

Data availability

The data that support the findings of this study are openly available at <https://data.mendeley.com/datasets/w7fgy7jvs8/2>, reference number²³.

Received: 9 December 2024; Accepted: 10 February 2025

Published online: 20 February 2025

References

- Kasapbaşı, A., Elbushra, A. E. A., Omar, A. H. & Yilmaz, A. DeepASLR: A CNN based human computer interface for American Sign Language recognition for hearing-impaired individuals. *Comput. Methods Progr. Biomed. Update* 2, 100048 (2022).
- Saleh, Y. & Issa, G. Arabic sign language recognition through deep neural networks fine-tuning (2020).
- Narayanan, V., Nithya, P. & Sathya, M. Effective lung cancer detection using deep learning network. *J. Cogn. Hum Comput Interact.* 2, 15–5 (2023).
- Wen, F., Zhang, Z., He, T. & Lee, C. AI enabled sign language recognition and VR space bidirectional communication using triboelectric smart glove. *Nat. Commun.* 12(1), 5378 (2021).
- Saraladeve, L. et al. A multiclass attack classification Framework for IoT using Hybrid Deep Learning Model. *J. Cybersecur. Inform. Manag.* 15(1) (2025).
- Hermawati, S. & Pieri, K. Assistive technologies for severe and profound hearing loss: Beyond hearing aids and implants. *Assist. Technol.* (2020).
- Zhang, Z. et al. Artificial intelligence-enabled sensing technologies in the 5G/internet of things era: from virtual reality/augmented reality to the digital twin. *Adv. Intell. Syst.* 4(7), 2100228 (2022).
- Katoch, S., Singh, V. & Tiwary, U. S. Indian Sign Language recognition system using SURF with SVM and CNN. *Array* 14, 100141 (2022).
- Papastratis, I., Chatzikonstantinou, C., Konstantinidis, D., Dimitropoulos, K. & Daras, P. Artificial intelligence technologies for sign language. *Sensors* 21(17), 5843 (2021).
- Daniel, E., Kathiresan, V. & Sindhu, P. Real Time Sign Recognition using YOLOv8 object detection algorithm for Malayalam sign Language. *Fusion: Pract. Appl.*, (1), 135–35 (2025).
- Akhila Thejaswi, R., Rai, B. S. & Pakkala, P. G. R. Graph data science-driven framework to aid auditory and speech impaired individuals by accelerating sign image analysis and knowledge relearning through deep learning technique. *Int. J. Syst. Assur. Eng. Manag.*, 1–24 (2024).
- Latif, G. et al. Learning at Your Fingertips: An Innovative IoT-Based AI-Powered Braille Learning System. *Appl. Syst. Innov.* 6(5), 91 (2023).
- Shwany, Z. A. A. et al. April. A Real-Time Hand Sign Language Recognition System for Threatening Situations Using Deep Learning. In 2024 12th International Symposium on Digital Forensics and Security (ISDFS), 1–6 (IEEE, 2024).
- Lakshmi, G., Pranav, S. & Deepak, S. December. Empowering Deaf and Mute Children through Computer Vision. In 2023 Intelligent Computing and Control for Engineering and Business Systems (ICCEBS), 1–6 (IEEE, 2023).

15. Akdag, A. & Baykan, O. K. Enhancing signer-independent recognition of isolated sign language through advanced deep learning techniques and feature fusion. *Electronics* **13**(7), 1188 (2024).
16. Namoun, A. et al. Service selection using an ensemble meta-learning classifier for students with disabilities. *Multimodal Technol. Interact.* **7**(5), 42 (2023).
17. Parveen, M. S., Keerthana, R. G., Shanmathi, S. & Shajitha, M. April. Sign Language Detection Using Open Cv. In *2024 International Conference on Communication, Computing and Internet of Things (IC3IoT)*, 1–5 (IEEE, 2024).
18. Faisal, M. et al. Enabling two-way communication of Deaf using Saudi sign Language. *IEEE Access* **11**, 135423–135434 (2023).
19. Li, J., Wang, T., Ma, C., Lin, Y. & Yan, Q. A file archival integrity check method based on the BiLSTM + CNN model and deep learning. *Egypt. Inform. J.* **29**, 100597 (2025).
20. Ghadi, Y. Y. et al. A and Machine learning solution for the security of wireless sensor network. *IEEE Access* (2024).
21. Zholshiyeva, L., Zhukabayeva, T., Baumuratova, D. & Serek, A. Design of QazSL sign Language Recognition System for physically impaired individuals. *J. Rob. Control (JRC)* **6** (1), 191–201 (2025).
22. Ghadi, Y. Y. et al. Integration of federated learning with IoT for smart cities applications, challenges, and solutions. *PeerJ Comput. Sci.* **9**, e1657 (2023).
23. Thakur, D., Dangi, S. & Lalwani, P. A novel hybrid deep learning approach with GWO–WOA optimization technique for human activity recognition. *Biomed. Signal Process. Control* **99**, 106870 (2025).
24. Mazhar, T. et al. Analysis of challenges and solutions of IoT in smart grids using AI and machine learning techniques: A review. *Electronics*, **12**(1), 242 (2023).
25. John, J. & Deshpande, S. Intelligent hybrid hand gesture recognition system using deep recurrent neural network with chaos game optimization. *J. Exp. Theor. Artif. Intell.* **37** (1), 75–94 (2025).
26. Renjith, S., Manazhy, R. & Suresh, M. S. February. Recognition of Sign Language Using Hybrid CNN–RNN Model. In *International Conference On Innovative Computing And Communication*, 17–28 (Springer Nature Singapore, 2024).
27. Paul, S. K. et al. IoT-Based Real-Time Medical-Related Human Activity Recognition Using Skeletons and Multi-Stage Deep Learning for Healthcare. Preprint at <https://arXiv.org/2501.07039> (2025).
28. Palanisamy, M., Mohanraj, R., Karthikeyan, A. & Mohanraj, E. December. SIGNEASE: AI-Driven American Sign Language Interpretation System. In *2024 International Conference on IoT Based Control Networks and Intelligent Systems (ICICNIS)*, 1670–1675 (IEEE, 2024).
29. Awarayi, N. S., Twum, F., Hayfron-Acquah, J. B. & Owusu-Agyemang, K. A bilateral filtering-based image enhancement for Alzheimer's disease classification using CNN. *PLoS One* **19** (4), e0302358 (2024).
30. Li, Y. et al. Tea leaf disease and insect identification based on improved MobileNetV3. *Front. Plant Sci.* **15**, 1459292 (2024).
31. Zhu, M. et al. High-precision monitoring and prediction of mining area surface subsidence using SBAS-InSAR and CNN-BiGRU-attention model. *Sci. Rep.* **14**(1), 28968 (2024).
32. Al-qaness, M. A., Ewees, A. A., Aseeri, A. O. & Abd Elaziz, M. Wind power forecasting using optimized LSTM by attraction-repulsion optimization algorithm. *Ain Shams Eng. J.*, 103150 (2024).
33. <https://data.mendeley.com/datasets/w7fgv7jvs8/2>
34. Luqman, H. An efficient two-stream network for isolated sign language recognition using accumulative video motion. *IEEE Access* **10**, 93785–93798 (2022).
35. Al Abdullah, B., Amoudi, G. & Alghamdi, H. Advancements in sign Language Recognition: a Comprehensive Review and Future prospects. *IEEE Access* (2024).

Acknowledgements

The authors extend their appreciation to the King Salman center For Disability Research for funding this work through Research Group no KSRG-2024- 062.

Author contributions

Conceptualization: Mashael Maashi, Huda G. Iskandar and Mohammed Rizwanullah, Data curation and Formal analysis: Mashael Maashi, Huda G. Iskandar and Mohammed Rizwanullah, Investigation and Methodology: Mashael Maashi, Huda G. Iskandar and Mohammed Rizwanullah, Project administration and Resources: Mashael Maashi, Writing—original draft: Mashael Maashi, Huda G. Iskandar and Mohammed Rizwanullah, Validation and Visualization: Mashael Maashi, Huda G. Iskandar and Mohammed Rizwanullah, Writing—review and editing, Mashael Maashi, Huda G. Iskandar and Mohammed Rizwanullah, All authors have read and agreed to the published version of the manuscript.

Funding

The authors thank the King Salman Center for Disability Research for funding this work through Research Group no KSRG-2024- 062.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025