A Real-Time Sign Language Recognition Framework Using Deep Learning and Internet of Things

Lama Al Khuzayem[1,*], Soukeina Elhassen[1]

[1]Department of Computer Science, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia

Emails: lalkhuzayem@kau.edu.sa; selhassen@stu.kau.edu.sa

Abstract

Sign language is a vital communication mean for hearing-impaired individuals, combining manual gestures with non-manual signs like facial expressions and body movements, often requiring both hands and sequential actions. Recently, an automatic Sign Language Recognition (SLR) has gained increasing attention, with Machine Learning and Deep Learning systems achieving competitive performance. While convolutional neural network has been widely employed owing to their effectiveness in image-based recognition tasks, existing methods, however, often struggle with efficiency, adaptability, and real-time deployment. This paper proposes an Internet of Things-Integrated Deep Learning Model for Real-Time SLR to enhance the communication among individuals with hearing-impairment and non-signers. The framework employs IoT-based wearable sensors for capturing hand and finger movements, followed by Sobel filtering for noise reduction. MobileNetV3 is applied for lightweight feature extraction, while a Variational AutoEncoder enables robust sign detection. To further improve performance, an Improved Sparrow Search Algorithm is introduced for hyperparameter tuning, constituting the novelty of this work. Experimental results show that the proposed framework achieves an outstanding accuracy of 99.05% when compared to state-of-the-art systems, validating its robustness and effectiveness for real-time SLR applications. Future work will explore large-scale deployment and multi-language adaptability.

Keywords: Internet of Things; Deep Learning; Sign Language Recognition; Hearing-Impaired; Improved Sparrow Search Algorithm

1. Introduction

In today's developing society, effective communication remains a fundamental necessity, encompassing individuals with hearing-impairments. It is estimated that there are more than 300 different sign languages worldwide, used by 72 million deaf or hard-of-hearing persons [1]. In this regard, Sign Language (SL) functions as a collective mode of expression, establishing a shared linguistic foundation for people in this community [2]. Over sustained growth, the developments in SL translation approaches and software have become transformative, immensely enriching the living standards for those who depend on SL as their main source of communication [3]. Particularly, the implementation of Sign Language Recognition (SLR) technology has recently experienced significant growth, specifically in the vibrant deaf community, efficiently enabling seamless interaction among its members [4].

Additionally, SLR applications have evolved in parallel to the Internet of Things (IoT) technology, which incorporates many devices, networks, and sensors [5, 6]. The authors of [7] survey the combination of IoT with real-time SL translators, emphasizing the role of networked devices in enhancing accessibility. Alsharif and Ilyas [8] discuss IoT applications in healthcare for hearing-impaired individuals, highlighting the potential for real-time

DOI: https://doi.org/10.54216/FPA.210225
Received: April 17, 2025 Revised: June 22, 2025 Accepted: August 30, 2025

assistive communication systems. SLs exhibit substantial regional variation and complexity, with different countries using distinct systems that often incorporate non-manual cues, sequential gestures, or two-handed movements [9]. Such diversity poses significant challenges for building accurate and scalable SLR systems. To address these challenges, investigators have increasingly focused on advanced SLR techniques [10], with recent efforts leveraging Machine Learning (ML) and Deep Learning (DL) systems. For instance, Zhang and Jiang [11] delivered a comprehensive review of state-of-the-art SLR approaches that exploit DL techniques, highlighting the diverse models and datasets used to improve recognition accuracy.

Multiple datasets are generated owing to multiple aspects such as type of images (Depth or RGB), regional differences, and more. SL varies from one area to another just similar to spoken languages [12]. Additionally, the images utilized for recognition methods relies on cameras producing depth or RGB images. Moreover, the methods applied by diverse investigators that underline the primary gesture recognition methods. Each study attempts to improve accuracy by developing new methods [13]. Presently, no system can deal with each condition with higher precision. Previous studies have often focused on Convolutional Neural Network (CNN)-based approaches due to their strong performance in image classification, experimenting with different parameters and architectures for SLR tasks [14].

Most present studies on SLR studies utilizing DL rely on CNN, Recurrent Neural Networks (RNNs), or hybrid methods, which attain good accuracy but struggle with challenges such as variations in hand shape, background complexity, motion dynamics, and signer diversity. Furthermore, numerous works concentrate on static databases and lack real-time flexibility, while problems like scalability, robustness across diverse environments, and integration with IoT for real-world deployment remain underexplored. This generates a research gap for developing more generalized, efficient, and real-time frameworks for SLR.

This paper presents an IoT-Integrated DL Model for Real-Time SLR (IoTDLM-RTSLR), designed to support individuals with hearing impairments by tackling the limitations of existing SLR systems. Unlike prior approaches that rely mainly on CNN-based classifiers or static datasets, the proposed framework uniquely integrates IoT-based wearable sensing, lightweight feature extraction, probabilistic modelling, and adaptive hyperparameter optimization. This combination enables robust, noise-resilient, and real-time detection of sign gestures, ensuring higher adaptability and scalability for real-world deployment.

The main contribution of this work lies in the design and development of IoTDLM-RTSLR framework, which incorporates four key components: (i) Sobel Filter (SF) for noise-robust pre-processing, (ii) MobileNetV3 for lightweight and efficient feature extraction, (iii) a Variational AutoEncoder (VAE) for robust gesture modelling and detection, and (iv) an Improved Sparrow Search Algorithm (ISSA) for adaptive hyperparameter optimization. By combining these elements, the system addresses critical challenges in SLR, including noise sensitivity, computational efficiency, generalization to unseen gestures, and parameter tuning. Experimental results show that IoTDLM-RTSLR achieves an outstanding result of 99.05% accuracy and strong robustness in real-time recognition tasks when compared to advanced approaches.

2. Related Work

Recent studies have explored a wide range of technologies to advance SLR and assistive systems for hearing-impaired individuals. For instance, Sharma et al. [15] proposed a TinyML-based solution employing low-cost wearable IoT devices to perform SLR. They deployed a lightweight Deep Neural Network (DNN) on edge devices that collects time-series motion sensor data for SLR, achieving 87.18% accuracy and ~10 ms inference time. The study also addressed the challenge of limited labelled data through Transfer Learning (TL) strategies, thereby improving training efficiency.

Building on IoT integration, Maashi et al. [16] proposed using hybrid DL models, MobileNetV3 combined with the Adaptive Residual Optimization Algorithm (AROA), to develop an aassistive communication tool for the hearing-impaired, achieving 99.19% accuracy. Goyal and Basavarajappa [17] designed and developed a wearable IoT-based haptic-aided gadget to assist interaction with the hearing-impaired. The device facilitates initiation of communication between hearing-impaired individuals and their family members by employing vibration and switch modules. As a proof-of-concept, Bluetooth Low Energy (BLE)-based wireless gadgets were tested and realized. A key benefit of proposed assistive device is its simplicity of use, low cost, and ease of recognition, thereby providing greater convenience for hearing-impaired users.

DOI: https://doi.org/10.54216/FPA.210225
Received: April 17, 2025 Revised: June 22, 2025 Accepted: August 30, 2025

Shirisha et al. [18] proposed an innovative method for transforming SL gestures into text, by employing progressive ML models and image processing to precisely translate and interpret SL the method employs a CNN classifier to process captured hand gestures and convert them into text. Similarly, Buttar et al. [19] developed a hybrid DL model utilizing You Only Look Once version 6 (YOLOv6) and the Long Short-Term Memory (LSTM) uniting static and dynamic sign detection, addressing the challenges of sequential gestures in SLR.

Xu et al. [20] proposed RF-CSign, a framework designed to achieve higher precision in cross-domain recognition and SLR. The system employs Radio Frequency Identification (RFID) to gather signals, and then denoise them during pre-processing. To mitigate the challenge of overfitting, RF-CSign integrates a Normalization-based Attention Module (NAM). The framework demonstrates improved precision in cross-domain scenarios by utilizing a migration learning approach.

Shaban and Elsheweikh [21] developed an Android-based tool for Arabic Sign Language (ArSL) and American Sign Language (ASL) recognition. Their design comprises a Sensory Smart Glove System (SSGSys), an IoT-enabled wearable intended for automated SLR. The authors of study [22], proposed a hand glove that includes embedded sensors to capture the user's physiological and motion-related signals. Using the collected dataset, ML models, such as K-Nearest Neighbors (KNN), were applied to classify users' physiological state as normal or abnormal. Compared with existing RFID Tag Reader–based approaches, the proposed solution offers several advantages.

Revathi et al. [23] presented a method that extracts spectrograms from speech signals and generates CNN–based models for classification. The evaluation compared performance across different acoustic features, including spectrogram, CNN, Gammatonegram, and Melspectrogram representations. Furthermore, speech intelligibility for hearing-impaired individuals was further enhanced through the application of the Phase Spectrum Compensation (PSC) model.

Al Khuzayem et al. [24] introduced Efhamni, a DL-based application for Saudi SLR. The method employs CNN with Bidirectional Long-Short Term Memory (BiLSTM) for processing visual Saudi Sign Language (SSL) gestures, achieving high accuracy. The system is designed to assist hearing-impaired individuals by providing real-time translation, enhancing communication accessibility. Similarly, Maashi et al. [25] presented a Harris Hawk Optimization (HHO)-based DL model for SLR, achieving 98.95% accuracy by integrating ResNet-152 feature extraction with Bi-LSTM classification. Elhassen and Al Khuzayem [26] for isolated Saudi SLR, achieving superior performance over CNNs with 97.50% accuracy on the dataset of KSU-ArSL, explored transformer-based models.

Although these studies, summarized in Table 1, demonstrate significant advances across IoT, wearable devices, hybrid deep learning, and transformer-based models, persistent challenges remain in terms of scalability, robustness across diverse environments, and real-time adaptability. These limitations motivate the development of our IoT-Integrated Deep Learning Framework (IoTDLM-RTSLR), which addresses these gaps through optimized preprocessing, lightweight feature extraction, robust detection, and hyperparameter tuning.

Table 1: SLR Related Literature Summary and Comparative Analysis

Research Study	Goal	Used Models	Accuracy
Sharma et al. [15]	To present a wearable IoT device that collects data using motion sensors, offers real-time sign identification, transmits the prediction to the cloud platform, and reduces the necessity for labeled data.	DNN and Deep TL	87.18%
Maashi et al. [16]	To create an IoT-driven assistive SLR tool using hybrid DL models for the hearing impaired.	SACHI-SLRHDL (MobileNetV3 + CNN-BiGRU-A + AROA)	99.19% on Indian SL Dataset (images)

DOI: https://doi.org/10.54216/FPA.210225
Received: April 17, 2025 Revised: June 22, 2025 Accepted: August 30, 2025

Goyal and Basavarajappa [17]	To design and develop an IoT-based wearable haptic-aided device to assist hearing-impaired individuals.	IoT Network	NA
Shirisha et al. [18]	To examine an innovative method for converting SL gestures into text.	CNN	96.7%
Buttar et al. [19]	To propose a novel model for recognizing words from gestures. The complexity relies in developing a technique for continuous sign recognition that is signer-independent.	YOLOv6 and LSTM	96%
Xu et al. [20]	To achieve superior performance in SLR and cross-domain recognition. The model employs large-kernel convolution to reduce complexity and capture long-range correlations, thereby improving recognition.	TL	99.17%, 96.67% and 97.50%
Shaban and Elsheweikh [21]	To develop intelligent systems intended mainly for automated SLR. Furthermore, the method acts as an auxiliary tool to learn SL, improving communication efficiency between signers and non-signers.	SSGSys and Mobile Augmented Sign Language Learning System (MASLL-Sys)	98.42% and 98.22%
Senthilnayaki et al. [22]	To propose an embedded device that performs specific SLR tasks using IoT technologies.	KNN	NA
Revathi et al. [23]	To present a speech command recognition system that translates speech into text for hands-free device control and improves accessibility for both hearing-impaired and normal users.	CNN	95%, 98%, and 99%
Al Khuzayem et al. [24]	To develop Efhamni, a deep learning-based application for recognizing SSL for the ease of communicating with the hearing-impaired community.	CNN with BiLSTM	94.46% on ArSL dataset
Maashi et al. [25]	To develop a novel DL model for automatic sign language recognition, enhancing communication for hearing impaired individuals by integrating advanced feature extraction and optimization techniques.	HHODLM-SLR (ResNet-152 + Bi-LSTM + HHO)	98.95% on SL Dataset
Elhassen and Al Khuzayem [26]	To establish a benchmark for isolated Saudi Sign Language recognition using transformer-based models, improving accessibility for the Saudi deaf community.	Transformer-based (Swin Transformer)	97.50% on KSU-ArSL (RGB video clips)

DOI: https://doi.org/10.54216/FPA.210225
Received: April 17, 2025 Revised: June 22, 2025 Accepted: August 30, 2025

3. Proposed Methodology

In this study, we propose the IoTDLM-RTSLR model to aid hearing-impaired individuals. The model enhances communication for individuals with hearing impairments using robust SLR methods. The IoTDLM-RTSLR framework is portrayed in Figure 1. The figure shows that the model includes IoT devices for collecting images, pre-processing using the SF, MobileNetv3 feature extractor, VAE classifier, and ISSA-based parameter optimizer. In the following subsections, we explain each phase in detail.

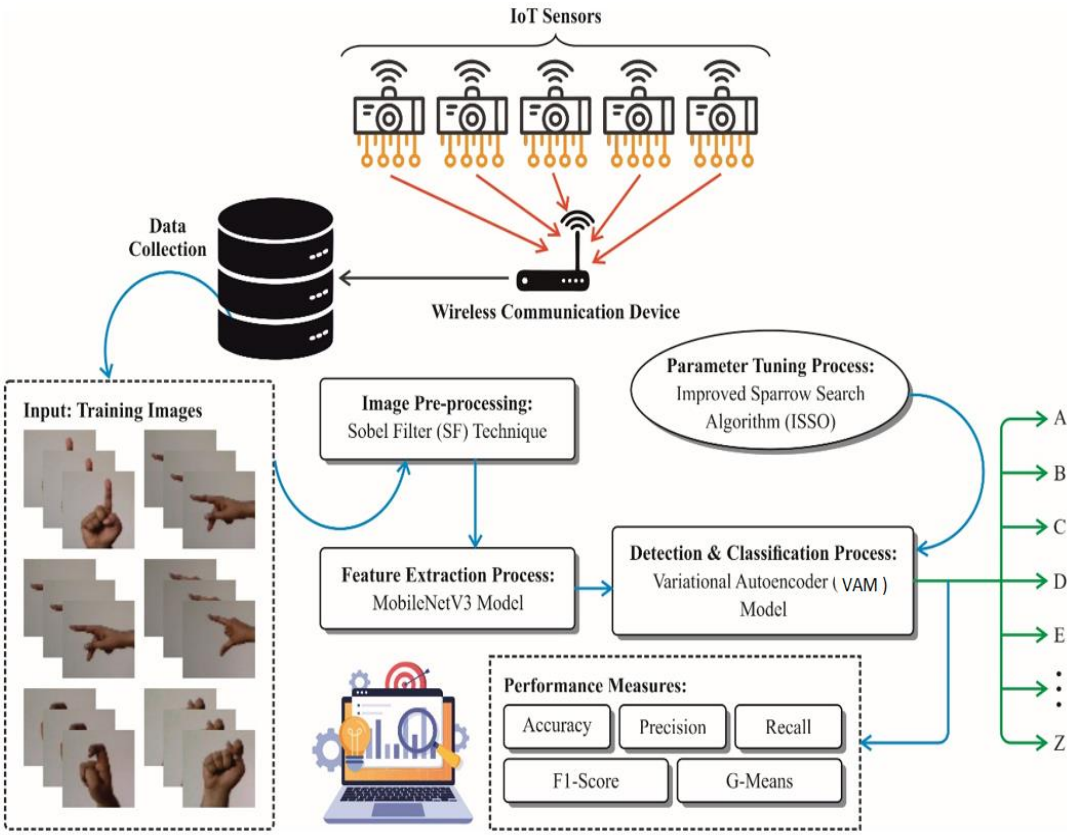

Figure 1. Workflow of IoTDLM-RTSLR Model

A. Noise Reduction

Initially, the SF is employed to eliminate the noise from the data. Sobel is selected because it reduces minor noise through gradient approximation while preserving significant structural details, making it appropriate for denoising tasks that require edge preservation. Unlike the median filter or Gaussian blur, which smooth images globally, the SF suppresses noise while retaining essential edge information. The SF is a widely used edge detection method in image processing and is particularly effective for SLR systems designed to support individuals with hearing impairments [27]. It works by computing the image intensity gradient, highlighting areas with sharp intensity changes that correspond to edges. This makes it ideal for detecting the outlines of fingers and hands, which are fundamental for identifying gestures in SL. By using vertical and horizontal kernels, the SF captures edge information in both directions, ensuring robust feature extraction. Its efficiency, simplicity, and ability to preserve critical gesture information make it a valuable tool in pre-processing for DL techniques. Integrating the SF improves the model's capability to accurately classify and recognize SL gestures, supporting reliable real-time communication solutions for persons with hearing impairments. The Sobel operator typically uses a 3×3 kernel.

B. MobileNetV3 Feature Extractor

The feature extraction procedure is performed by MobileNetV3 model. MobileNetV3 was preferred over heavier CNNs such as ResNet because it provides a lightweight architecture with enhanced depth-wise separable convolutions, making it extremely effective for real-time applications. It achieves a good balance among accuracy

DOI: https://doi.org/10.54216/FPA.210225
Received: April 17, 2025 Revised: June 22, 2025 Accepted: August 30, 2025

and computational cost, which is vital for deployment on IoT and mobile devices. The MobileNetV3 network in our model demonstrates superior feature extraction performance compared to other related approaches [28]. It is a lightweight network capable of capturing richer information from high-dimensional features while alleviating gradient-related issues produced by shortcut networks. It incorporates a lightweight attention mechanism, linear bottleneck inverted residual blocks from MobileNetV2, depth-wise separable convolutions from MobileNetV1, and the h-swish activation function as a replacement for the earlier swish function. The enhanced capabilities of MobileNetV3 have been widely applied in detection, segmentation, and classification tasks. For comparison, Wang et al. [29] introduced YOLOv7, an advanced lightweight model for real-time recognition, which serves as a benchmark for evaluating MobileNetV3's efficiency in our SLR system.

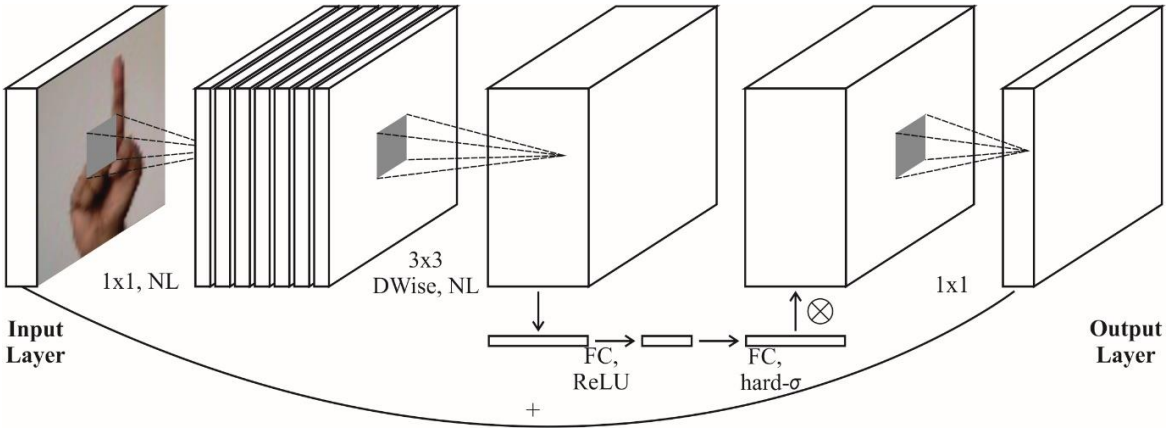

Figure 2. Framework of MobileNetV3

The process starts with global average pooling to compress spatial information. This is followed by an excitation stage comprising two fully connected (FC) layers: the first decreases the dimensionality to one-quarter of channel size and applies an activation of ReLU, while the second one restores the channel dimension using h-swish function. The SE block then generates a set of channel weights, which are applied element-wise to feature mapping of size W × H × C, thereby strengthening informative channels and suppressing less related ones. The overall architecture of MobileNetV3 is shown in Figure 2.

C. SL Recognition and Classification using VAE

The VAE model is employed for sign language detection and classification. MobileNetV3 is first employed for extracting higher-level gesture features from input images, which are fed into the VAE to learn a compact latent representation for strong classification. VAE is preferred over standard CNN methods because it not only achieves classification but also models the fundamental data distribution, which ultimately makes the model robust against noise, variations, and unseen gestures, which traditional CNNs struggle to handle. Classification is achieved by mapping input signs into the latent space and applying a classifier (e.g., Softmax or MLP) on the latent vectors. In some cases, latent space clustering can also be employed to group similar gestures for detection.

The methodology of this study is depend upon the core concept of the VAE approach, which aims to attain efficient optimization and data generation through a probabilistic generative method [30]. Kezar et al. [31] explored VAE-based representation learning for isolated signs, showing improved generalization for out-of-vocabulary gestures, which aligns with our approach. VAE primely optimizes the generation method by maximizing the log-likelihood probability $P(x)$ of an observed data. Let x represent data sample, and the aim is to learn the latent representation z such that the model can generate data similar to x under the condition of provided z. Initially, consider the log-likelihood of the detected sample x:

$$\log p(x) = \int q\,(z|x) \log \frac{p(x,z)}{q(z,x)}\, dz \tag{1}$$

DOI: https://doi.org/10.54216/FPA.210225
Received: April 17, 2025 Revised: June 22, 2025 Accepted: August 30, 2025

As a sample, $q(x|z)$ refers to variational distributions that are applied to estimate the posterior distribution $p(z|x)$. By presenting the variational distribution $q(x|z)$, the above equation can be decomposed into dual portions:

$$\log p(x) = E_{q(z|x)}[\log p(x|z)] - D_{KL}(q(z|x)\|p(z)) \quad (2)$$

The first term corresponds to the reconstruction error, which represents the probability of generating data presented by latent variable z; the second term is Kullback-Leibler (KL) divergence, which measures the distance between an estimated distribution and the true prior distribution $p(z)$. By minimizing the sum of reconstruction error and KL divergence, the generative performance of the VAE is enhanced.

The encoder contains three Fully Connected (FC) layers of 512, 256, and 128 neurons, followed by $ReLU$ activation functions. The latent space has a dimensionality of 64. The decoder mirrors the encoder with layers of 128, 256, and 512 neurons, respectively, utilizing $ReLU$ activations. The decoder's output layer utilizes a sigmoid activation to produce normalized output.

To achieve the diversity and continuity of generation, this study also introduces the Gaussian distribution theory of latent variable z. Assuming the prior distribution of latent variable z represents standard normal distribution $p(z) = N(0, I)$, its conditional distribution $p(x)$ can additionally be presumed to be a Gaussian distribution, namely,

$$q(z|x) = N(\mu(x), \sigma^2(x)) \quad (3)$$

whereas, $\mu(x)$ and $\sigma(x)$ means outputs of the parameterized network. During model training, the reparameterization trick is applied to ensure gradient transferability. The latent variable z can be stated as:

$$z = \mu(x) + \sigma(x) \cdot \varepsilon \quad (4)$$

where $\varepsilon \sim N(0, I)$ represents standard normal distribution noise. Through this transformation, noise is added to the latent variable generation method to guarantee the generated diversity.

In the optimization procedure, the objective function is the sum of reconstruction error and the KL divergence:

$$L = E_{q(z|x)}[\log p(x|z)] - D_{KL}(q(z|x)\|p(z)) \quad (5)$$

The parameters are updated using the gradient descent approach to meet user-specific requirements while retaining a degree of authenticity.

To further enhance the model's adaptive design, this paper offers a feedback mechanism based on user's behavior. The interactive data is fed to the model as an additional input, influencing its adjustment and generation. On the assumption that the feedback is f, the generated is stated as $p(x|z, f)$. By modeling this conditional distribution, the model is dynamically optimized based on the habits and preferences of the users. The last model's objective function is provided as:

$$L = E_{q(z|x)}[\log p(x|z, f)] - D_{KL}(q(z|x)\|p(z)) \quad (6)$$

This enhanced design allows the model to be dynamically optimized according to user preferences and habits. The proposed framework is inspired by the work of Zhang et al. [29].

D. ISSA-based Parameter Tuning

Finally, the ISSA optimizes the hyperparameter values and resulting in enhanced classification performance. This study introduces ISSA to improve hyperparameter selection in oversampling and address the challenges associated with parameter tuning [32]. Fan et al. [33] demonstrated the efficacy of a hybrid SSA for optimizing hyperparameters in deep learning models, supporting its application in the VAE tuning process. The Tent Chaotic Mapping (TCM) improves the SSA by exploiting its unpredictability features, uniform distribution, and irreducibility. Data produced by the TCM function are used as an initial positional values for the sparrows, which aids to maintain search diversity and facilitates effective exploration. This enhancement contributes to faster convergence and greater SSA global search capability.

The Tent Chaotic Mapping (TCM) function, also denoted to as Tent mapping function, produces uniformly distributed chaotic sequences in a defined interval and is characterized by a higher iteration speed. The Tent map formula is expressed in Equation (7):

DOI: https://doi.org/10.54216/FPA.210225
Received: April 17, 2025 Revised: June 22, 2025 Accepted: August 30, 2025

$$x_{i+1} = \begin{cases} 2x & 0 \le x_j \le 1/2 \\ 2\,(1 - xi) & 1/2 \le xi \le 1 \end{cases} \quad (7)$$

Because Tent chaotic iterations may converge to small unstable periodic points, random variables are incorporated into the initial TCM function. Furthermore, the Tent mapping function is subjected to a Bernoulli shift transformation for enhancing the descriptive control, as well as to improve the stability and convergence speed of the method. The modified formulation is expressed in Equation (8):

$$x_{i+1} = (2x_i) mod 1 + rand(0,1) \times \frac{1}{N} \quad (8)$$

In the ISSA model, the discoverer sparrows are more likely to seek food, leading to higher fitness values and moving closer to the global optimum. In each search iteration, the discoverer updates its location according to Equation (9):

$$X_{i,j}(t+1) = \begin{cases} X_{i,j}(t) \cdot \exp\left(-\frac{i}{\alpha \cdot t_{\max}}\right) & if R < ST \\ X_{i,j}(t) + Q \cdot L & if R \ge ST \end{cases} \quad (9)$$

Where $X_{i,j}$ represents the $i-th$ sparrow in $j-th$ size; t denotes a present iteration, and $t_{\max}$ indicates the maximal iteration count; α is randomly generated of $[0,1]$; R is in an interval of $(0,1)$, which represents the sparrow's warning value, triggering a chirping call upon encountering danger; ST is in the interval of [0.5, 1] that represents the threshold at which the model will transfer to a safer location for foraging; Q is a randomly generated number; and L denotes matrix with dimensions lxd with all components equal to 1.

When the finder completes a comprehensive hunt for appropriate food resources, the followers travel toward the finder's place. When the foraging try fails, the sparrows relocate to unexplored areas, thus increasing exploration and overall fitness. The behaviour is modeled in Equation (10):

$$X_{i,j}(t+1) = \begin{cases} Q \cdot \exp\left(\frac{X_w - X_{i,j}(t)}{i^2}\right) & if\ i > \frac{n}{2} \\ X_{i,j}(t) + \left|X_{i,j}(t) - X_p(t)\right| \cdot A^* \cdot L & otherwise \end{cases} \quad (10)$$

Here X_p characterizes the location where the finder contains an additional food, namely, the optimum fitness value; X_w characterizes the location by the poor fitness value; A refers to matrix with dimensions lxd and components 1 or -1 and fulfills relation $A^* = A^T(AA^T)^{-1}$. The position of the sparrow population is further updated by Equation (11):

$$X_{i,j}(t+1) = \begin{cases} X_b(t) + \beta \cdot \left|X_{i,j}(t) - X_b(t)\right| & if f_i \ne f_b \\ X_{i,j}(t) + K \cdot \frac{\left|X_{i,i}(t) - X_w(t)\right|}{(f_i - f_w) + \varepsilon} & if f_i = f_b \end{cases} \quad (11)$$

The fitness evaluation is the critical factor driving ISSA's performance. Hyperparameter selection is encoded as candidate solutions, with fitness assessed based on classification accuracy. The fitness function was defined below:

$$Fitness = \max(P) \quad (12)$$

Where

$$P = \frac{TP}{TP + FP} \quad (13)$$

Here, TP is True and FP is False Positive value, respectively.

4. Experimental Validation

The performance analysis of IoTDLM-RTSLR technique is validated utilizing the Sign language Gesture Image Dataset [34]. The dataset consists of 39,000 samples across 26 class labels, as shown in Table 2. Figure 3 illustrates sample images.

DOI: https://doi.org/10.54216/FPA.210225
Received: April 17, 2025 Revised: June 22, 2025 Accepted: August 30, 2025

Table 2: Details of the Dataset

Sign	Label	Number of Samples	Sign	Label	Number of Samples
A	L-1	1500	N	L-14	1500
B	L-2	1500	O	L-15	1500
C	L-3	1500	P	L-16	1500
D	L-4	1500	Q	L-17	1500
E	L-5	1500	R	L-18	1500
F	L-6	1500	S	L-19	1500
G	L-7	1500	T	L-20	1500
H	L-8	1500	U	L-21	1500
I	L-9	1500	V	L-22	1500
J	L-10	1500	W	L-23	1500
K	L-11	1500	X	L-24	1500
L	L-12	1500	Y	L-25	1500
M	L-13	1500	Z	L-26	1500
		Total Number of Samples			39000

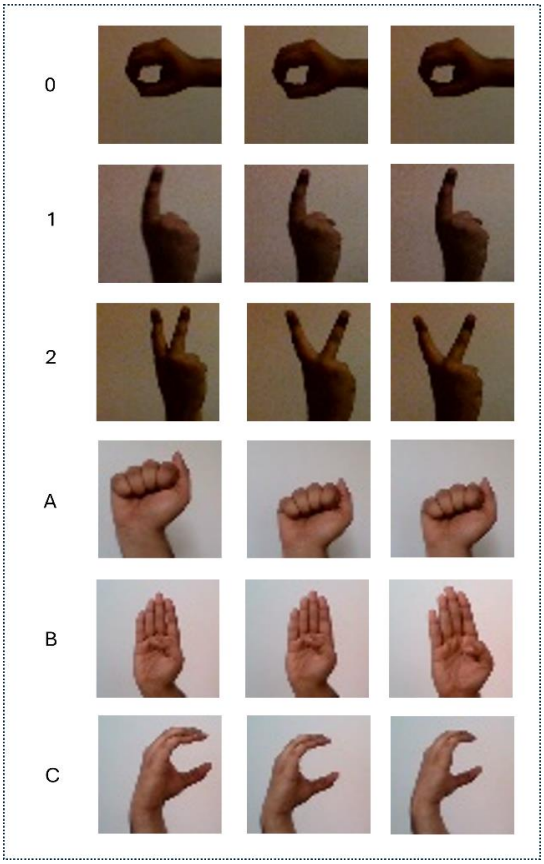

Figure 3. Sample Images of the Dataset

DOI: https://doi.org/10.54216/FPA.210225
Received: April 17, 2025 Revised: June 22, 2025 Accepted: August 30, 2025

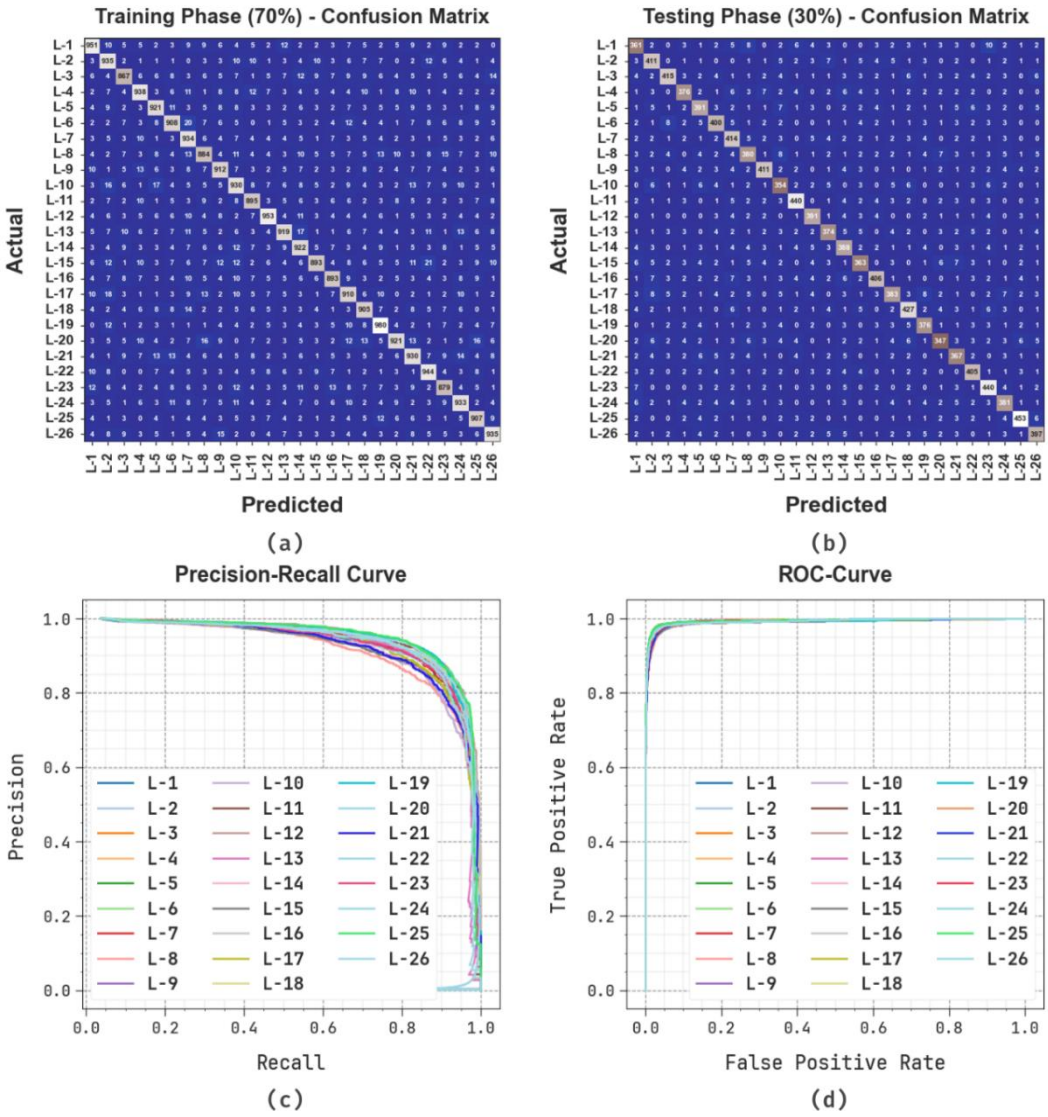

Figure 4. IoTDLM-RTSLR Classification Analysis (a-b) 70% TRPH and 30% TSPH Confusion Matrices, (c-d) Curves of PR and ROC

Figure 4 presents the classification analysis of the IoTDLM-RTSLR model. Specifically, Figures 4a–4b depict the confusion matrices under 70% TRPH and 30% TSPH, confirming accurate classification and identification across all classes. Figure 4c displays the Precision–Recall (PR) curves, highlighting consistently strong performance across categories. Figure 4d illustrates the Receiver Operating Characteristic (ROC) curves, demonstrating robust results with high values for ROC across the various class labels.

Table 3 and Figure 5 illustrate the sign language recognition of IoTDLM-RTSLR algorithm under 70% and 30%. The performances indicate that the IoTDLM-RTSLR system accurately recognized the samples. Under 70% TRPH, the IoTDLM-RTSLR method achieved average $accu_y$,$prec_n$,$reca_l$, $F1_{score}$ and G_{Means} of 99.04%, 87.58%, 87.55%, 87.54%, and 87.55%, respectively. Under 30% TSPH, the IoTDLM-RTSLR model achieved average $accu_y$,$prec_n$,$reca_l$, $F1_{score}$ and G_{Means} of 99.05%, 87.59%, 87.55%, 87.55%, and 87.56%, respectively.

DOI: https://doi.org/10.54216/FPA.210225
Received: April 17, 2025 Revised: June 22, 2025 Accepted: August 30, 2025

Table 3: IoTDLM-RTSLR Results for 70% and 30%

Class Labels	$Accu_y$	$Prec_n$	$Reca_l$	$F1_{score}$	G_{Means}
TRPH (70%)					
L-1	99.13	89.21	88.63	88.92	88.92
L-2	99.01	85.00	89.99	87.42	87.46
L-3	99.01	88.47	84.75	86.57	86.59
L-4	99.08	88.41	87.91	88.16	88.16
L-5	99.09	88.30	87.88	88.09	88.09
L-6	99.02	87.64	86.72	87.18	87.18
L-7	98.96	83.69	90.24	86.84	86.90
L-8	98.86	85.99	84.11	85.04	85.05
L-9	99.06	87.52	87.78	87.65	87.65
L-10	98.78	84.16	85.48	84.82	84.82
L-11	99.13	88.18	88.35	88.26	88.26
L-12	99.19	90.33	88.98	89.65	89.65
L-13	99.01	88.45	85.97	87.19	87.20
L-14	98.95	85.69	87.48	86.57	86.58
L-15	99.03	91.12	83.38	87.08	87.17
L-16	99.08	88.77	86.61	87.68	87.68
L-17	98.94	85.29	87.33	86.30	86.30
L-18	99.07	86.77	88.64	87.69	87.70
L-19	99.18	88.53	91.08	89.78	89.79
L-20	99.06	90.21	85.52	87.80	87.83
L-21	98.90	86.03	86.19	86.11	86.11
L-22	99.15	88.22	89.90	89.06	89.06
L-23	99.02	87.20	86.35	86.77	86.77
L-24	99.06	87.85	87.94	87.89	87.89
L-25	99.19	88.23	90.16	89.18	89.19
L-26	99.10	87.79	88.96	88.37	88.38
Average	99.04	87.58	87.55	87.54	87.55

DOI: https://doi.org/10.54216/FPA.210225
Received: April 17, 2025 Revised: June 22, 2025 Accepted: August 30, 2025

TSPH (30%)					
L-1	98.89	84.94	84.54	84.74	84.74
L-2	98.99	85.80	89.15	87.45	87.46
L-3	99.13	91.21	87.00	89.06	89.08
L-4	99.09	88.26	86.84	87.54	87.55
L-5	98.95	86.31	86.50	86.41	86.41
L-6	99.04	87.15	88.30	87.72	87.72
L-7	98.91	84.49	89.03	86.70	86.73
L-8	98.85	85.39	84.63	85.01	85.01
L-9	99.18	89.93	89.15	89.54	89.54
L-10	98.91	83.69	85.92	84.79	84.80
L-11	99.21	90.53	90.35	90.44	90.44
L-12	99.23	88.26	91.14	89.68	89.69
L-13	99.08	88.00	86.77	87.38	87.39
L-14	98.89	84.35	87.00	85.65	85.66
L-15	98.97	86.84	84.62	85.71	85.72
L-16	99.09	90.42	86.57	88.45	88.47
L-17	98.91	88.05	83.62	85.78	85.81
L-18	99.08	88.41	89.14	88.77	88.77
L-19	99.16	88.26	88.68	88.47	88.47
L-20	98.96	88.30	82.03	85.05	85.11
L-21	99.00	85.35	87.17	86.25	86.26
L-22	99.20	89.21	90.00	89.60	89.60
L-23	99.08	86.96	91.29	89.07	89.10
L-24	99.00	86.59	86.79	86.69	86.69
L-25	99.34	92.64	91.70	92.17	92.17
L-26	99.09	88.03	88.42	88.22	88.22
Average	99.05	87.59	87.55	87.55	87.56

DOI: https://doi.org/10.54216/FPA.210225
Received: April 17, 2025 Revised: June 22, 2025 Accepted: August 30, 2025

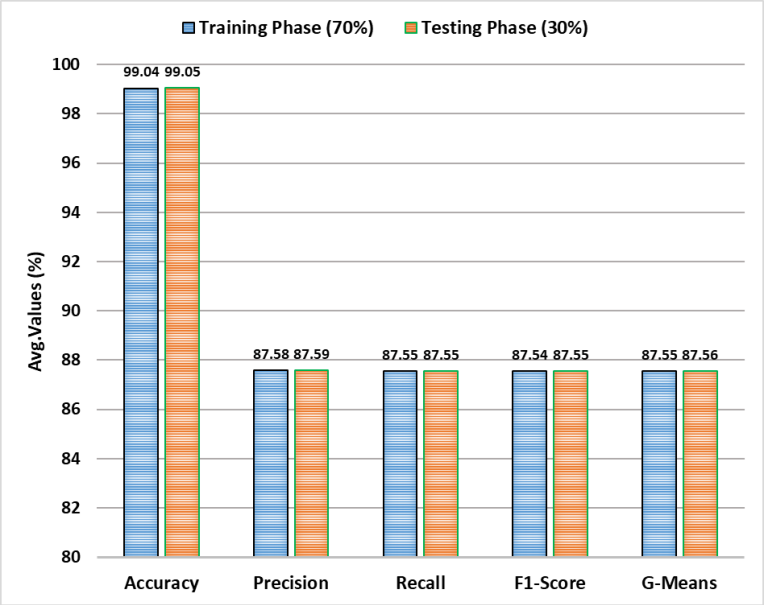

Figure 5. Average Values of the IoTDLM-RTSLR Model under 70% and 30%

Figure 6 shows the IoTDLM-RTSLR training (TRA) $accu_y$ and validation (VAL) $accu_y$. The accuracy $accu_y$ values are computed over 0-25 epochs. The figure highlights that both TRA and VAL accuracy exhibit an increasing trend, indicating the effectiveness of IoTDLM-RTSLR approach with higher performance across multiple repetitions. In addition, the TRA and VAL accuracy values stay close across epochs, indicating balanced fitting and presenting superior results of IoTDLM-RTSLR technique, ensuring reliable predictions on unseen data.

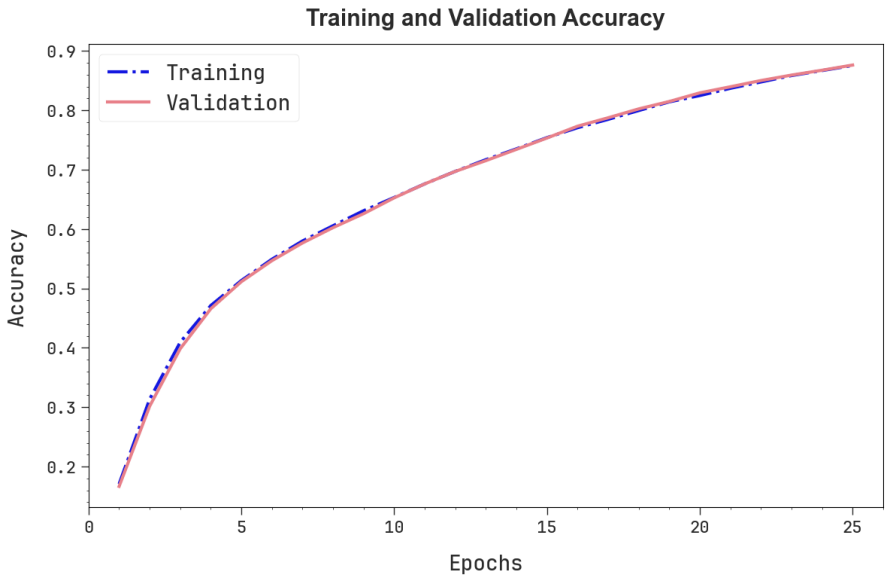

Figure 6. The IoTDLM-RTSLR Model's $Accu_y$ Curve

Figure 7 shows the IoTDLM-RTSLR model's training loss (TRALOS) and validation loss (VALLOS) curves, which are calculated over 0-25 epochs. The figure illustrates a decreasing trend in the values of both, indicating the proficiency of the IoTDLM-RTSLR model in balancing data fitting and generalization. The steady decline indicates that the IoTDLM-RTSLR system achieves optimal performance while progressively refining its computational outcomes.

DOI: https://doi.org/10.54216/FPA.210225
Received: April 17, 2025 Revised: June 22, 2025 Accepted: August 30, 2025

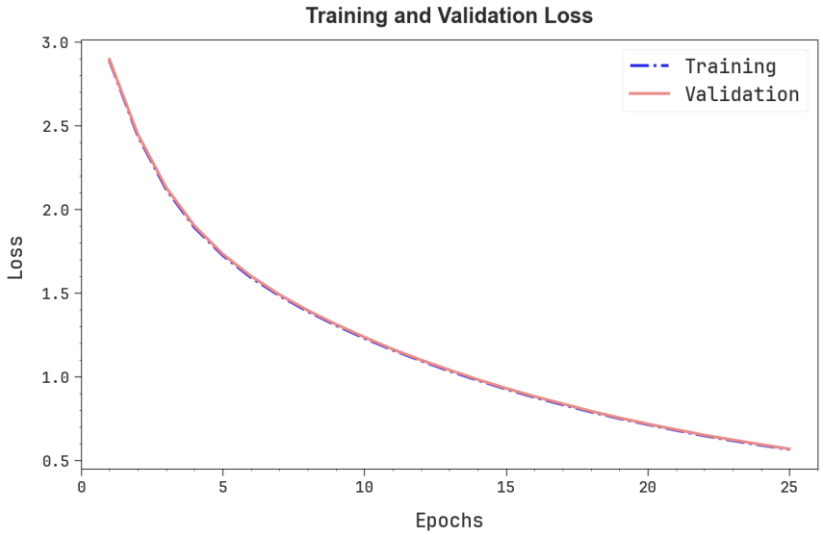

Figure 7. Loss Curve of IoTDLM-RTSLR Model

Table 4 presents a detailed ablation study of the IoTDLM-RTSLR technique. The proposed model without ISSA model has demonstrated lower performances with $accu_y$ of 97.66%, $prec_n$ of 86.47%, $reca_l$ of 86.35%, and $F1_{score}$ of 86.02%, respectively. However, the proposed IoTDLM-RTSLR methodology (with ISSA) has depicted a higher performance with $accu_y$ of 99.05%, $prec_n$ of 87.59%, $reca_l$ of 87.55%, and $F1_{score}$ of 87.55%, correspondingly.

Table 4: Ablation Study of the IoTDLM-RTSLR Model

Approach	$Accu_y$	$Prec_n$	$Reca_l$	$F1_{score}$
Proposed Without ISSA	97.66	86.47	86.35	86.02
IoTDLM-RTSLR (With ISSA)	99.05	87.59	87.55	87.55

Table 5 and Figure 8 presents a comparative results of the IoTDLM-RTSLR model against existing methodologies [35-41]. The results show that the Random Forest [35], DNN-Squeezenet [36], CNN Classifier [37], RBM [38], ML-CNN [39], Conv LSTM [40], and VGG16 [41] models have stated poorer performance. Meanwhile, VGG16 algorithm has accomplished quicker solutions. Meanwhile, the IoTDLM-RTSLR technique indicated superior performance with improved $prec_n$, $reca_l$, $accu_y$, and $F1_{score}$ of 87.59%, 87.55%, 99.05%, and 87.55%, correspondingly.

Table 5: A Comparative Study of IoTDLM-RTSLR with State-of-the-art Techniques

Approach	$Accu_y$	$Prec_n$	$Reca_l$	$F1_{score}$
Random Forest [35]	90.00	83.05	75.21	80.90
DNN-Squeezenet [36]	83.28	81.66	80.38	80.36
CNN Classifier [37]	93.00	76.07	85.19	86.64
RBM Method [38]	98.13	82.39	77.78	84.01
ML-CNN [39]	88.00	87.00	86.00	87.50
Conv LSTM [40]	87.00	85.00	87.00	86.00
VGG16 [41]	86.00	84.00	85.00	83.00
IoTDLM-RTSLR	99.05	87.59	87.55	87.55

DOI: https://doi.org/10.54216/FPA.210225
Received: April 17, 2025 Revised: June 22, 2025 Accepted: August 30, 2025

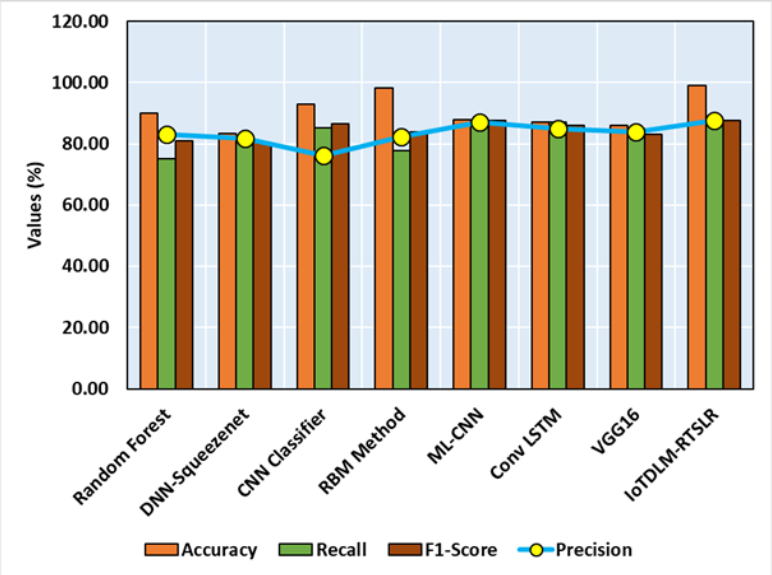

Figure 8. Comparative Analysis of IoTDLM-RTSLR Model with Existing Approaches

Table 6 and Figure 9 report computational time. The IoTDLM-RTSLR achieved the shortest runtime of 1.22sec, outperforming RF, DNN-Squeezenet, CNN, RBM, ML-CNN, Conv LSTM, and VGG16, which required between 3.91 and 9.13 seconds.

Table 6: Time Outcome of the IoTDLM-RTSLR Model with Recent Methods

Approach	Time (sec)
Random Forest [35]	4.10
DNN-Squeezenet [36]	8.65
CNN Classifier [37]	6.38
RBM Method [38]	6.54
ML-CNN [39]	4.99
Conv LSTM [40]	3.91
VGG16 [41]	9.13
IoTDLM-41RTSLR	1.22

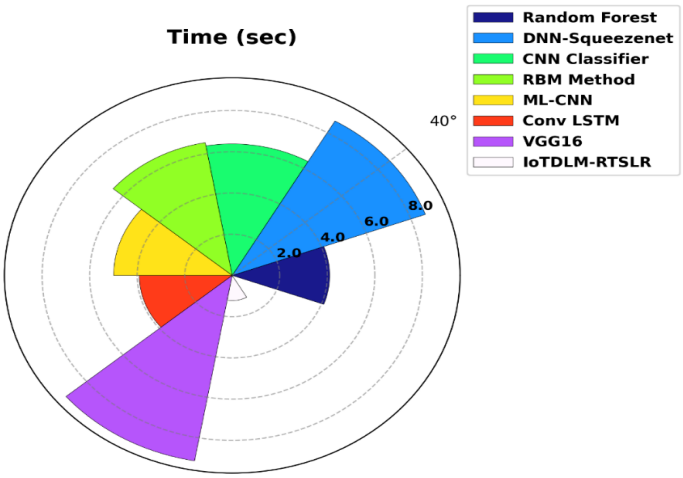

Figure 9. Time outcome of IoTDLM-RTSLR Model against Recent Models

DOI: https://doi.org/10.54216/FPA.210225
Received: April 17, 2025 Revised: June 22, 2025 Accepted: August 30, 2025

5. Conclusion

In this paper, an IoTDLM-RTSLR framework to support individuals with hearing impairments was proposed. The model enhances communication through robust SLR. First, the SF is employed in the preprocessing stage to eliminate input noise. Next, MobileNetV3 extracts features efficiently. The VAE is then applied for detection and classification of gestures. Finally, ISSA optimizes the hyperparameters of the VAE, resulting in higher classification performance. An extensive set of experiments confirmed the advanced results of IoTDLM-RTSLR, achieving an accuracy of 99.05% along with strong precision, recall, and F1-score metrics. These results validate the robustness and reliability of the method for real-time sign language recognition, supporting accessible and effective communication solutions for the hearing-impaired individuals. For future work, this framework can be extended to larger and more diverse sign language databases, real-world IoT environments, and multi-modal inputs (e.g., facial expression and body posture). Additionally, expanding the system to handle continuous sign language sequences and multiple sign languages will further enhance its applicability and global impact.

Funding: "This research received no external funding"

Conflicts of Interest: "The authors declare no conflict of interest."

References

[1] National Geographic Society. "Sign Language." *National Geographic Resource Library*. Accessed: [Date Accessed]. [Online]. Available: https://education.nationalgeographic.org/resource/sign-language/#

[2] V. Jain, A. Jain, A. Chauhan, S. S. Kotla, and A. Gautam, "American sign language recognition using support vector machine and convolutional neural network," *Int. J. Inf. Technol.*, vol. 13, pp. 1193–1200, 2021.

[3] K. Bantupalli and Y. Xie, "American Sign Language Recognition using Deep Learning and Computer Vision," in *Proc. IEEE Int. Conf. Big Data*, 2018, pp. 4896–4899.

[4] S. Hollier and S. Abou-Zahra, "Internet of Things (IoT) as assistive technology: Potential applications in tertiary education," in *Proc. 15th Int. Web for All Conf.*, 2018, pp. 1-4.

[5] J. Hou et al., "Signspeaker: A real-time, high-precision smartwatch-based sign language translator," in *Proc. 25th Annu. Int. Conf. Mobile Comput. Netw.*, 2019, pp. 1-15.

[6] V. Bhatnagar, R. Chandra, and V. Jain, "IoT based alert system for visually impaired persons," in *Proc. Int. Conf. Emerging Technol. Comput. Eng. (ICETCE)*, 2019, pp. 216-223.

[7] M. Papatsimouli, P. Sarigiannidis, and G. F. Fragulis, "A survey of advancements in real-time sign language translators: integration with IoT technology," *Technologies*, vol. 11, no. 4, p. 83, 2023.

[8] B. Alsharif and M. Ilyas, "Internet of things technologies in healthcare for people with hearing impairments," in *IoT and Big Data Technologies for Health Care*. Cham: Springer, 2022, pp. 299–308.

[9] T. W. Chong and B. G. Lee, "American sign language recognition using leap motion controller with machine learning approach," *Sensors*, vol. 18, no. 10, p. 3554, 2018.

[10] Y. Ma, G. Zhou, S. Wang, H. Zhao, and W. Jung, "SignFi: Sign language recognition using WiFi," *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.*, vol. 2, no. 1, pp. 1-21, 2018.

[11] Y. Zhang and X. Jiang, "Recent Advances on Deep Learning for Sign Language Recognition," *Comput. Model. Eng. Sci.*, vol. 139, no. 3, 2024.

[12] I. Papastratis et al., "Artificial intelligence technologies for sign language," *Sensors*, vol. 21, no. 17, p. 5843, 2021.

[13] M. A. Ahmed et al., "Based on wearable sensory device in 3D-printed humanoid: A new real-time sign language recognition system," *Measurement*, vol. 168, p. 108431, 2021.

[14] A. A. Alhussan, M. M. Eid, and W. H. Lim, "Advancing Communication for the Deaf: A Convolutional Model for Arabic Sign Language Recognition," *Full Length Article*, vol. 5, no. 1, pp. 38-48, 2023.

[15] S. Sharma, R. Gupta, and A. Kumar, "A TinyML solution for an IoT-based communication device for hearing impaired," *Expert Syst. Appl.*, vol. 246, p. 123147, 2024.

[16] A. Almjally and W. S. Almukadi, "Deep computer vision with artificial intelligence based sign language recognition to assist hearing and speech-impaired individuals," *Sci. Rep.*, vol. 15, p. 32268, 2025.

[17] M. Goyal and G. Basavarajappa, "A Wearable IoT Based Assistive Device to Aid Communication With Hearing Impaired," in *Proc. IEEE Microw. Antennas Propag. Conf. (MAPCON)*, 2023, pp. 1-4.

[18] N. Shirisha et al., "Hand Talk: Sign Language to Text Converter using CNN," in *Proc. 8th Int. Conf. I-SMAC*, 2024, pp. 1654-1659.

DOI: https://doi.org/10.54216/FPA.210225
Received: April 17, 2025 Revised: June 22, 2025 Accepted: August 30, 2025

[19] A. M. Buttar et al., "Deep learning in sign language recognition: a hybrid approach for the recognition of static and dynamic signs," *Mathematics*, vol. 11, no. 17, p. 3729, 2023.

[20] H. Xu, Y. Zhang, Z. Yang, H. Yan, and X. Wang, "RF-CSign: A Chinese Sign Language Recognition System Based on Large Kernel Convolution and Normalization-Based Attention," *IEEE Access*, vol. 11, pp. 133767-133780, 2023.

[21] S. A. Shaban and D. L. Elsheweikh, "An Intelligent Android System for Automatic Sign Language Recognition and Learning," *J. Adv. Inf. Technol.*, vol. 15, no. 8, 2024.

[22] B. Senthilnayaki et al., "Enhanced Health Monitoring Using IoT-Embedded Smart Glove and Machine Learning," in *Proc. Int. Conf. Innov. Sustain. Comput. Technol. (CISCT)*, 2023, pp. 1-5.

[23] A. Revathi, N. Sasikaladevi, D. Arunprasanth, and N. Raju, "Raspberry Pi-based robust speech command recognition for normal and hearing-impaired (HI)," *Multimedia Tools Appl.*, vol. 83, no. 17, pp. 51589-51613, 2024.

[24] L. Al Khuzayem et al., "Efhamni: A deep learning-based Saudi sign language recognition application," *Sensors*, vol. 24, no. 10, p. 3112, 2024.

[25] M. Maashi, H. G. Iskandar, and M. Rizwanullah, "IoT-driven smart assistive communication system for the hearing impaired with hybrid deep learning models for sign language recognition," *Sci. Rep.*, vol. 15, p. 6192, 2025.

[26] S. Elhassen and L. Al Khuzayem, "Bridging information science and deep learning: Transformer models for isolated Saudi Sign Language recognition," *Appl. Math. Inf. Sci.*, vol. 19, no. 6, pp. 1345–1357, 2025.

[27] D. Sharifrazi et al., "Fusion of convolution neural network, support vector machine and Sobel filter for accurate detection of COVID-19 patients using X-ray images," *Biomed. Signal Process. Control*, vol. 68, p. 102622, 2021.

[28] J. Di, W. Guo, J. Liu, L. Ren, and J. Lian, "AMMNet: A multimodal medical image fusion method based on an attention mechanism and MobileNetV3," *Biomed. Signal Process. Control*, vol. 96, p. 106561, 2024.

[29] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2023, pp. 7464-7475.

[30] R. Zhang, S. Wang, T. Xie, S. Duan, and M. Chen, "Dynamic User Interface Generation for Enhanced Human-Computer Interaction Using Variational Autoencoders," *arXiv preprint arXiv: 2412.14521*, 2024.

[31] L. Kezar, Z. Sehyr, and J. Thomason, "Phonological Representation Learning for Isolated Signs Improves Out-of-Vocabulary Generalization," *arXiv preprint arXiv: 2509.04745*, 2025.

[32] S. Ye, B. Da, L. Qi, H. Xiao, and S. Li, "Condition Monitoring of Marine Diesel Lubrication System Based on an Optimized Random Singular Value Decomposition Model," *Machines*, vol. 13, no. 1, p. 7, 2025.

[33] Y. Fan et al., "A hybrid sparrow search algorithm of the hyperparameter optimization in deep learning," *Mathematics*, vol. 10, no. 16, p. 3019, 2022.

[34] A. Khanak, "Sign Language Gesture Images Dataset," *Kaggle*, 2022. [Online]. Available: https://www.kaggle.com/datasets/ahmedkhanak1995/sign-language-gesture-images-dataset

[35] C. Dong, M. C. Leu, and Z. Yin, "American sign language alphabet recognition using microsoft kinect," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2015, pp. 44–52.

[36] N. Kasukurthi, B. Rokad, S. Bidani, and D. Dennisan, "American Sign Language Alphabet Recognition using Deep Learning," *arXiv preprint arXiv: 1905.05487*, 2019.

[37] W. Tao, M. C. Leu, and Z. Yin, "American Sign Language alphabet recognition using Convolutional Neural Networks with multiview augmentation and inference fusion," *Eng. Appl. Artif. Intell.*, vol. 76, pp. 202–213, 2018.

[38] R. Rastgoo, K. Kiani, and S. Escalera, "Multi-modal deep hand sign language recognition in still images using restricted Boltzmann machine," *Entropy*, vol. 20, no. 11, p. 809, 2018.

[39] N. Aslam, K. Abid, and S. Munir, "Robot assist sign language recognition for hearing impaired persons using deep learning," *VAWKUM Trans. Comput. Sci.*, vol. 11, no. 1, pp. 245–267, 2023.

[40] A. Baihan, A. I. Alutaibi, and M. Alshehri, "Sign language recognition using modified deep learning network and hybrid optimization: a hybrid optimizer (HO) based optimized CNNSa-LSTM approach," *Sci. Rep.*, vol. 14, p. 26111, 2024.

[41] S. Mohsin et al., "American sign language recognition based on transfer learning algorithms," *Int. J. Intell. Syst. Appl. Eng.*, vol. 12, no. 5s, pp. 390–399, 2024.

DOI: https://doi.org/10.54216/FPA.210225
Received: April 17, 2025 Revised: June 22, 2025 Accepted: August 30, 2025