SEMARAK ILMU PUBLISHING

Semarak Engineering Journal

Journal homepage:
https://semarakilmu.my/index.php/sej/index
ISSN: 3036-0145

The E-Learning Application in A Classroom for Deaf-Mute Students

Atyaf Hekmat[1,*], Hawraa Hasan Abbas[2], Zeinab A.Alhusein Almulla[3], Blessing Olamide Taiwo[4]

1 Directorate General of Education in Karbala Province, Karbala, Iraq
2 College of Information Technology Engineering, Al-Zahraa University for Women 56001 Karbala, Iraq.& Department of Electrical and Electronic Engineering, University of Kerbala Karbala 56001, Iraq
3 College of Education, Al-Zahraa University for Women 56001 Karbala, Iraq
4 Mining Engineering, Engineering Technology, Federal University of Technology Akure, Nigeria

ARTICLE INFO

Article history:
Received 20 July 2025
Received in revised form 29 November 2025
Accepted 10 December 2025
Available online 25 December 2025

Keywords:
Sign Language Recognition; Convolution Neural Network (CNN); Squeezenet

ABSTRACT

Usage of the E-learning environment has become one of the most important requirements of the current era, and the general trend of students today is to deal more passionately with these environments. Sign Language Recognition (SLR) is the most common way for deaf-mute to communicate among themselves or with others. Deaf-mute students need to communicate with their teachers or their colleagues. By e-learning application, this work is trying to translate the gestures of (SLR) to audible sound. Then use the voice to address Alexa. A widely used control method for smart environments is voice control, which is very difficult for the deaf community. Thus, the proposed system describes an automatic fast, and accurate system adopted for recognizing one-hand static gestures in real-time conditions. The base is a vision-based RGB image in different environments for American Sign Language (ASL). Deep Learning (DL) technique was used by Convolution Neural Networks (CNN) architecture in three models: Alexnet, Googlenet, and Squeezenet. The used dataset in the off-time implementation of two levels of complexity: RGB images of the uniform color background and RGB images with complex backgrounds. The best accuracy achieved for a complex background dataset of (6×150) image by the Squeezenet model, was 99.85% accuracy for off-time testing. Real-time experiments gave the best accuracy of 98.593 % by Squeezenet in 0.001sec recognition time. This achievement has been exploited for smart classrooms by providing a comment for 'Amazon Alexa', which is a powerful voice recognition system.

1. Introduction

The World Health Organization (WHO) published a survey indicating that in 2005, hearing loss impacted around 278 million individuals worldwide. Ten years later, that number had risen to 360 million, a 14% increment, and the number has continued steadily increasing since that time. In 2019, a total of 466 million people worldwide suffered from hearing loss, which accounted for 5% of the global population. This number consisted of 432 million adults (83%) and 34 million children (17%). The WHO has also predicted that, by 2050, this community may have more than doubled to around

[*] *Corresponding author.*
E-mail address: atyaf.h@s.uokerbala.edu.iq

https://doi.org/10.37934/sej.11.1.102121

900 million people [3]. Gesture languages are the primary means by which Deaf people communicate with each other, with various sign languages (SL) using a variety of hand motions, stances, and gestures having emerged globally, some of which also incorporate other communication features such as lip and eye movements. SLs are not identical across societies, varying in a similar manner to oral languages such as Arabic, Indian, or English. However, each society has units of communication that can be divided into static and dynamic gestures.

The lack of knowledge among hearing people of such languages results in a lack of comprehension among gesture users, as these signs or gestures may refer to letters, words, or full units of thought equivalent to sentences. E-learning offers a solution to this by providing automation systems that can translate a given SL into some form of oral or written language. It is an excellent way to increase the percentage of educated deaf-mute people. The deliberate use of networked information and communications technology for teaching and learning is known as e-learning. Many new systems are designed to assist deaf-mute people in accessing the world of learning and training [47].A multidisciplinary field of study called human-computer interaction (HCI) is devoted to the design of computer technology and, more specifically, to the interaction between people, or users, and computers [22].

There are three types of approaches to gesture recognition: sensor-based, vision-based, and hybrid. Many types of research have been carried out concerning each of these. Systems designed based on sensor techniques utilize devices such as smart gloves as input [14], while systems utilizing vision-based techniques often use a 2D camera to capture RGB images as input [11]. Other systems have combined devices such as the Microsoft Kinect, with its effective depth sensor, with RGB cameras to develop a hybrid technique(Bheda and Radpour 2017). Recently, there has been a call for hand gesture recognition systems not only to translate SLs as produced by Deaf people but also to support the development of smart environments for e-learning and medical applications based on most SLs speed and ease of understanding as compared to oral or written languages. Sounds may be changed and become unrecognizable when a speaker suffers from certain health issues, while the time taken to write a communication of a need or description will almost always exceed that required to perform the equivalent hand gesture [20].

Automated speech and language recognition (SLR) systems are often designed to detect and interpret generated signals, deciphering their significance in a manner comprehensible to individuals who speak different languages [8]. Nevertheless, this method is susceptible to numerous problems. HCI researchers have the primary difficulty of determining the appropriate data to train the applicable machines. Since signed hand gestures are a visually diversified form of communication, the data will differ based on the image itself and the surrounding surroundings. Additionally, there will be variations between static and dynamic hand gestures, as well as motions made with one or two hands. A system in Jayashree *et al*., [33] that employed static hand motions in real-time, even when there were complicated backgrounds present. Real-time status necessitates swift image processing, rendering static images more appropriate for this purpose. However, the intricacy of the backgrounds presented some challenges in the data. To tackle this, the author implemented pre-processing techniques, which further increased the processing and time requirements.

In Barbhuiya *et al*., [11], Abul Abbas employed a plain black background for the data items utilized in the system. This decision resulted in enhanced accuracy by precisely indicating that the information pertains solely to the gesture depicted in the image. Nevertheless, this setting was not the inherent one for users. Conversely, Pavlo utilized dynamic hand gestures in an actual system in reference [30]. Both systems were presented with uninterrupted streams of raw visual input that needed to be divided into individual frames, identified, and categorized into pre-established classifications, which added complexity to the task. A further significant obstacle involves choosing a suitable technique for extracting the characteristics of the image. Several research projects rely on feature extraction techniques (FE) to analyze images of gestures obtained from a camera or an input dataset. This step is crucial in the traditional machine learning (TML) technique, where features are extracted and data classes are manually classified [17]. Nevertheless, researchers frequently require pre-processing of the pictures prior to feature extraction, and the specific procedures for this phase differ depending on the desired characteristics. In Islam *et al*., [21], Md. Mohiminul implemented a system that computed the Euclidean distance between the two nearest fingertips of the hand motion image. This was followed by pre-processing steps, including shrinking the image and transferring it to a binary format. Ultimately, the system successfully identified the important characteristics following the application of a median filter to eliminate any interference caused by noise. Prior to feature extraction (FE), certain academics have attempted to separate the hand shape from the image's background. In their study, Hira Ansar [9] utilized various types of features, including geodesic distance, for extraction. On the other hand, Raghuveera *et al*., [36] opted to employ Microsoft Kinect for capturing motions, enabling easier segmentation in an untouched environment. From this, he derived three distinct types of features: Local Binary Pattern (LBP), Histograms of Oriented Gradients (HOG), and Speeded up Robust Features (SURF). In their study, Alzohairi and Reema conducted a comparison of four different feature extraction (FE) techniques: HOG, Edge Histogram Descriptor (EHD), LBP, Discrete Wavelet Texture Descriptor (DWT), and Gray-Level Co-occurrence Matrix (GLCM) [5]. The researchers in Ansar *et al.*, [9], Raghuveera *et al*., [36] and Alzohairi *et al*., [5] conducted comparisons between the findings of each characteristic and subsequently chose the most optimal one for their respective systems.

Following the process of feature extraction, these systems are still required to categorize the data. However, the precision of the outcomes may be contingent upon the type of classifier used. Hence, the third problem entails determining the classification methodology to be employed. Mirza *et al*., [40] employed Support Vector Machines (SVM), whereas Haque (Haque, Das, and Kaspy 2019) opted for the K-Nearest Neighbor (KNN) technique during the classification stage. Furthermore, other researchers have also encountered difficulties in determining the optimal approaches for data selection, whether it be through the TML techniques mentioned earlier or the emerging deep learning (DL) frameworks, which have shown promising outcomes. Therefore, these algorithms are frequently utilized, with the most prevalent one being CNN. The utilization of Convolutional Neural Networks (CNNs) in visual computing tasks has grown in recent years [7], primarily because the algorithm necessitates minimal pre-processing. In Barbhuiya *et al*., [11], Barbhuiya employed pre-trained neural network designs, namely AlexNet and VGG16, which consist of many layers, to interpret pertinent data. The initial layer of both networks represents the attributes of the input image. To ensure compatibility with the pre-trained networks, photos of varying sizes in the dataset must be scaled to satisfy the parameters of the input layer. Manual feature extraction is not performed in these modules. Instead, features are extracted within the hidden layers of the CNN network and subsequently classified in a separate layer. System in Shin *et al*., [41] used a CNN network In order to benefit from the long-range dependencies computation of the transformer and the local feature calculation of the CNN for sign language recognition, a convolution and transformer-based multi-branch network system for 77 label Korean SL dataset and 98.30% accuracy for the lab dataset

Kasukurthi and Nikhil Kasukurthi *et al.*, [25] employed a squeezenet, a variant of CNN, which possesses compact dimensions suitable for deployment on mobile devices. In order to strengthen the connection between the needs of the smart environment and addressing sign language (SL) requirements for the deaf community, certain researchers, like Luka in Kraljević *et al*., [27], have attempted to transform dynamic gestures into commands that can manipulate the smart home. They achieved this by utilizing a system consisting of a Jetson TX2 embedded device, developed in collaboration with NVIDIA, and a StereoLabs ZED M stereo camera. Examining the current body of literature reveals the numerous difficulties associated with this kind of task. These issues arise due to the diverse range of data entry types and the requirement to determine the most optimal approach. Various approaches were examined to ascertain the optimal choice in terms of speed, simplicity, suitability, and cost-effectiveness for all users requiring system implementation. The primary innovation of the proposed system is in its utilization of sign language gestures for controlling the system inside a smart classroom setting. Sign language gestures facilitate the administration of the system in an intelligent classroom setting. Utilize the RGB image using an intricate background, which accurately portrays the typical state of a student or user in real-time, subsequent to evaluating photographs with various shapes and situations. Selecting SqueezeNet as the optimal deep learning convolutional neural network (DL-CNN) system for real-time applications, in order to ensure a precise design in the DL system. The main contribution of the suggested method decreases memory usage by eliminating segmentation, intricate pre-processing, and the SqueezeNet feature. Therefore, section 2 of this paper offers the methodology and the proposed system of the work, while section 3 introduces outlines the results and its discussion. A conclusion and suggestions for future work are then offered in section 4.

2. Methodology

2.1 Materials

Types of datasets that used in the off-time phase:

2D static hand gestures shown as color images. These gestures featured simple black backgrounds, as shown in Figure1.

The second type featured very complex backgrounds: only six signs are included in this set, creating a mixture of images captured from a computer camera and those images from a dataset of the ASL alphabet. The term “Complex background” refers to images entered into the dataset without any amendments to the real background against which the photo is taken. These images are also of different sizes, with varying lighting forms, different skin tones, and different gesture positions. For each of the six sign investigated, 150 images were included: the number of test images is thus (6×150) =900, as illustrated in Figure2.

The final data type are a static hand gesture version of the full ASL alphabet (letters a - z plus a space). For every sign gesture, 50 images are included, so the full number of images is (27×50) =1,350, as illustrated in Figure3.

Fig. 1. RGB\2D Static Hand Gesture, ASL (bhuczak et al. 2011)

Fig. 2. Static Hand Gesture, ASL, complex background (“American Sign Language Detection Using PCA and LDA - File Exchange - MATLAB Central,” n.d.)

Fig. 3. Static Hand Gesture, full ASL alphabet (“ASL Alphabet | Kaggle” n.d.)

The PC specifications that used for the proposed systems, Intel(R) Core(TM) i7 3520MCPU@ 2.90 GHz, 8GB RAM, NVIDIA version391.25, were installed as a backend. Moreover, Matlab R2018b as an application program.

2.2 Methods

The methodology of this work is divided into two including the traditional machine learning approach and deep learning approach.

2.2.1 Traditional machine learning (TML) methodology

Traditional means the techniques that have been done for many years and are usually the basis for more cutting-edge machine learning. The speed and relative simplicity of TML is the key benefits. Some of these algorithms can be understood by humans. It making them crucial for failure analysis, Model development, and the identification of insights and statistical patterns. In this work, there are a test of some models by TML methodology is done to check if it is more suitable or the DL methodology. TML contains some steps beginning with preprocessing step, FE, and classification at last, as shown in Figure 4.

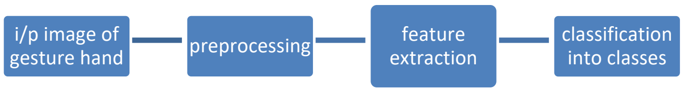

Fig. 4. Block diagram of TML

In this work, features like (HOG, LBP, and GLCM) are extracted, after make a preprocessing step by resizing the image and converting it to gray scale, at last the classification step by KNN is done to complete a TML system and obtain its results. This system was implemented to make a comparison with the DL system.

HOG features: it is based on occurrences of gradient orientation, thus providing edges and directions (Ullah, Mohammed, and Cheikh 2018). Images are resized to 64×64 pixels with conversion to grayscale is done for every image; the cell size chosen is [8 8] per image, giving a final matrix is of (k×1765).

LBP features: is a texture features and local descriptor of the image. It is based on the neighborhood for any given pixel (Suleiman 2018), All images are resized to 64×64 pixels, also the image is converted to its grayscale which means the value of the pixels is between (0~256). Then each pixel value in the image unless pixels on the image border, is compared with the values of its neighboring pixels. The LBP feature at last a feature vector of 1×236, making the full matrix k×236.

GLCM features: it is one of the statistical methods, which characterize the texture of an image by calculating how often pairs of pixels with specific values within a specified spatial relationship occur in an image [23]. When these features are utilized, resizing image to 64×64 pixels is required before they are converted to grayscale. The spatial relationship between pixels is then computed using a Gray-Level Co-occurrence Matrix. For every image, a feature vector of 1×4 is created, in the format (1-Contrast, 2-Correlation, 3-Energy, 4-Homogeneity). Finally, the features matrix that moves on to the classification stage takes the form k×4, where k is the number of images and (4) is the number of features.

2.2.2 Deep learning methodology

(DL) is an exciting field and an important branch of Artificial Intelligence (AI) because of its ability to extract meaningful patterns from the input data such as images. The discipline of DL has seen a rise in interest in the topic of computer programs analyzing images thanks to the quick growth of computer graphics. Image analysis is a group of processes performed on an image to enhance its appearance or extract relevant information [49]. Recent DL techniques can build a complete model that computes final classification labels from the original image pixels [45].

A deep neural network, which uses simple components running in parallel and is modelled after biological nerve systems, integrates several nonlinear processing layers (Mohd Akmal Masud et al. 2023). An input layer, a number of hidden levels and an output layer make up this structure. Nodes, or neurons, connect the layers, and each hidden layer uses the output of the preceding layer as its input.

DL using CNN has recently been a common tool used for SLR. CNN can be utilized in many algorithms like (Alex Net, VGG, Google Net, SeqeezeNet, YOLOnet,..). This technique can offer high accuracy, but it needs more computing time [11,32].

DL is done here with various different types of (CNNs). The choice of such architectures was favored because of the reduced need for an image pre-processing step, the hidden extracting features step characteristic, and the hidden classification features:

AlexNet: This CNN architecture, which was developed in 2012 with 61-million learnable parameters, implements max pooling to perform a down sampling operation. Usually it includes five-convolutional layers followed by five-relu layers. Then, there are five max-pooling layers, followed by two fully connected layers. AlexNet, thus, has a kernel of filters of sizes (11×11), (5×5), and (3×3) for its five convolutional layers [11].

GoogleNet: This network, built in 2015 with five million learnable parameters, is twelve times smaller than AlexNet but performs better. It is based on the Inception architecture, which increases the number of units in every layer by using parallel filters called the inception module. These filters of size (1×1), (3×3), and (5×5) in every convolution layer (Kim and O'Neill-Brown 2019). The main idea is to find the construction, which must be optimal locally and then repeat it spatially; this makes the layer more wide and lead to make the network wider rather than deeper.

SqueezeNet: One of the most important models in the field of image classification because of its distinguishing features and the accuracy of the results obtained when compared to other models.When SqueezeNet was created in 2016, it was reduced to less than one million weights. It is of 10-layers, then a convolution layer followed by 8-fire modules and another convolution layer at the end. This architecture achieves nearly the same accuracy as Alexnet, but with 50 times fewer parameters [48].

What distinguishes sequeeznet is the relatively small number of weights, which simplifies tasks and considerably speeds up the training process. The architecture of this net was proposed with several advantage by Wang *et al*., [48]. Sequeeznet follows the same framework as CNN in terms of the types of layers it has, and how it works in general. What makes it different is the policy of reducing and squeezing that is used with each layer or the overall optimization.

Starting with the convolution layers, which are the most important layers in all CNNs models, each has one or more matrices of filters that act as learnable weights. Filters are convolved with the input to compute the output, as a mathematical form they are slid across the input to compute the sum of the dot products between the filter and the input at every single position. Therefore, this computational process becomes more complicated and requires more time with the filters of large dimensions as if it is (5×5) or (3×3). In squeezenet filters in this layer are (1×1). For example, if compared to the filter (3×3) it is 1/9 less computationally [19].

Figure5 shows how to convolve of input image (5×5) with a filter (3×3) to give a map of (3×3). Getting the first element in the output is done by sliding the filter across the first (3×3) cells in the input: (1×1)+(1×0)+(1×1)+(0×0)+(1×1)+(1×0)+(0×1)+(0×0)+(1×1) = 4

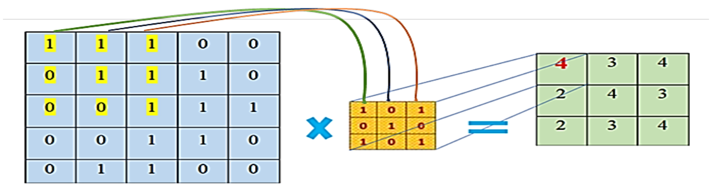

Fig. 5. Convolve a part of input image (5×5) by filter(3×3) to give map of(3×3)

Another reduction is for the input channels, which are reduced to (3×3) filters to get:

Total quantity of parameters in the layer = (number of input channels) × (number of filters) × (3×3)

For all the CNNs models, the output of the convolution layer is called 'Activation Map' [1]. Figure 6 shows the number of the Activation Maps for an input image of (32×32×3) which gives 6-activation maps of (28×28×3) as output if it is through 6-filters of (5×5). Squeezenet has also an improving design although the two minimization patterns are shown above, the design keep its big activation map design despite a late down sampling [19].

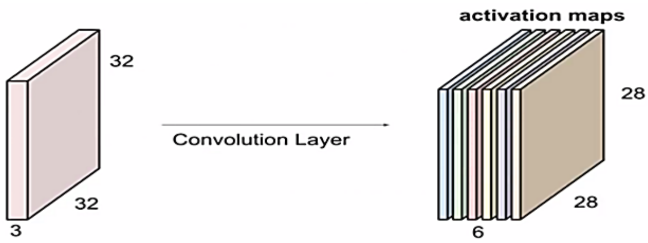

Fig. 6. Activation maps for (32×32×3) input image

A successful application of these three ideas, the squeezenet achieved them through a building block architecture called 'Fire module' which it is A squeeze convolution layer (that contains only 1x1 filters) feeding into an expanded layer with a combination of 1x1 and 3x3 convolution filters constitutes shown in Figure7.

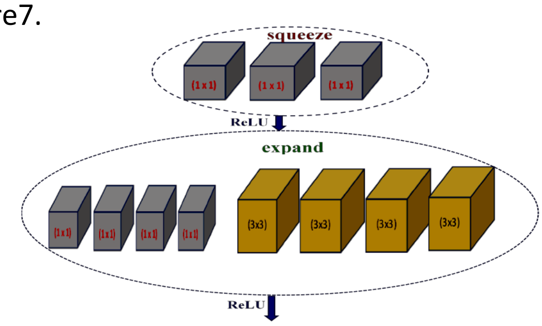

Fig. 7 Fire module structure in SqueezeNet model

A single convolution layer (conv1) is the beginning of the squeezenet design, then 8-Fire modules are followed that (Fire 2-9), at last another convolution layer (conv10). With each Fire module, the number of filters is gradually increased. Figure8 shows the details of this architecture

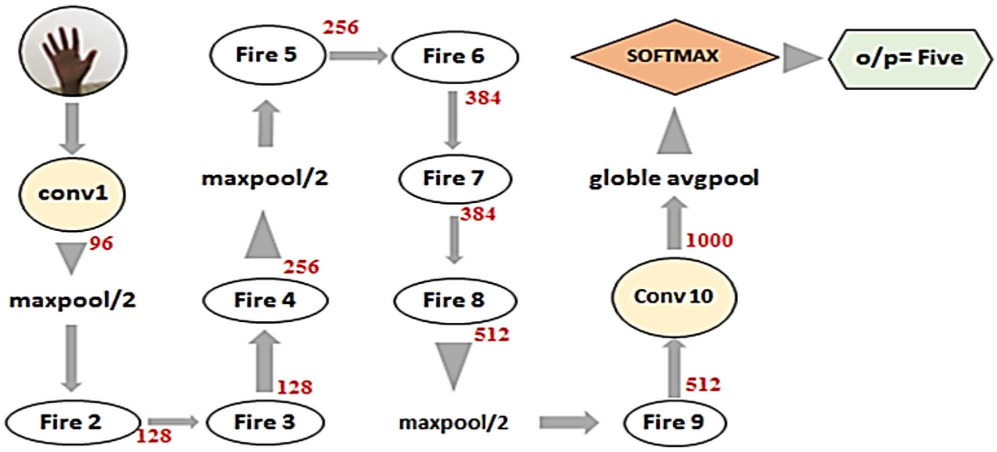

Fig.8. SqueezeNet architecture

At last the characteristics of the used networks was shown in Table 1 and Figure 9 shows the block diagram for our system, which is learned in the off-time and then subjected to several real-time tests.

Table 1
The characteristics of CNNs used

Network type	Input size in pixels	No. of layers in Matlab	Initial Learn Rate	Train to validation ratio
Alex Net	227×227×3	25	10^{-4}	70:30
Google Net	224×224×3	144	10^{-4}	70:30
Squeeze Net	227×227×3	68	10^{-4}	70:30

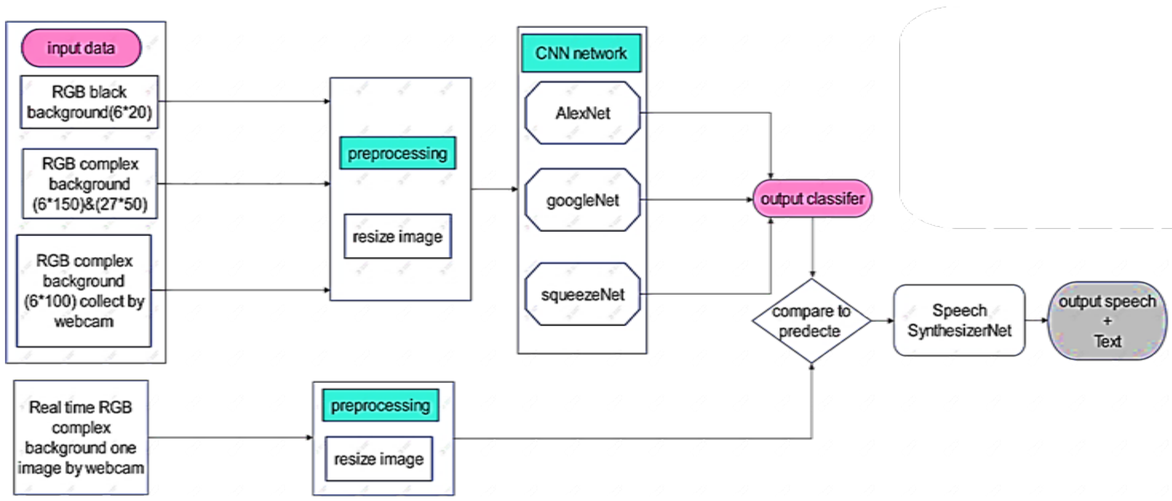

Fig. 9. Proposed DL architecture system

3. Results and Discussion

For the TML system, the results in terms of accuracy (Acc.) and the time required for the validation test are shown in Table 2 for (HOG, LBP, and GLCM) features using the KNN classifier.

Table2
Results for TML

Type of Feature	No. & type of signs	testing Time/sec	Acc. %
HOG	6×20 Black background	3.8	100
	6×150 Complex background	11.8	95
	(27×50)alphabet Complex Background	17.49	99.2
LBP	6×20 Black background	0.71	96.7
	6×150 Complex background	2.89	91.7
	(27×50) alphabet Complex Background	3.008	97.7%
GCLM	6×20 Black background	*3.5*	*88.3*
	6×150 Complex background	*0.75*	*81.8*
	(27×50) alphabet Complex Background	*1.22*	*90.2*

The validation results for the DL system in terms of accuracy (Acc.) and the time required for the (training and validation) time by the using of the 3-types of CNN technique in the off time are shown in Table 3. The best results appear with SqueezeNet.

Table3
Results for DL by three types of CNN

CNN type	No. of signs type image	NO. of layers	NO. of Epochs	Training time/min	Testing time/sec	Acc. %
AlexNet CNN	6*20 Black background	25	10	3.33	3.44	100
	6*150 Complex Background	25	20	127	11.44	98.50
GoogleNet CNN	6*20 Black background	144	5	6.31	6.64	99.8
	6*150 Complex Background	144	10	354	32.99	98.52
SqueezeNet CNN	6*20 Black background	68	5	4.13	3.12	100
	6*150 Complex Background	68	10	369	11.2	99.85
	(27*50) alphabet Complex Background	68	7	17hour	24.86	97.57

A comparison was offered between the three types of CNN in terms of:

- ✓ The year of invention.
- ✓ The number of parameters needed to reach the final decision necessary for the classification process.
- ✓ The ratio of minimizing the number of parameters when considering AlexNet as the reference (Ref.).
- ✓ The module that was implemented in each type.
- ✓ In addition, the memory size required by each model. All of that are shown in Table 4.

Table 4
Comparison between the three types of CNN

	CNN	Year of invention	No. of parameters in millions	Reduction ratio	Module implement	Memory size/ MB
1	**AlexNet**	2012	61	Ref.	Max pooling	202
2	**GoogleNet**	2015	5	12 time smaller	Inception module	21
3	**SqueezeNet**	2016	1	50 time smaller	Fire module	3

Hand gesture recognition is a crucial procedure that can have a significant impact on various applications in advanced fields. Furthermore, it enables enhanced communication for individuals with hearing impairments by employing artificial intelligence and computer visualization techniques. This study shows cases numerous endeavors to attain optimal outcomes in this field. Table 2 presents a diverse range of findings obtained by testing various sorts of feature extraction (FE) methods with the Histogram of Oriented Gradients (HOG) technique. Notably, the black backdrop yielded the highest achievement, reaching a perfect score of 100%. The experiment with the alphabet, conducted against a difficult background, yielded favorable outcomes with a 10-fold validation accuracy of 99.2%. By comparison, Alzohairi *et al.*, [5] attained a precision of 63.56% by employing a comparable technique. The utilization of LBP features for feature extraction showed that employing 10-fold cross-validation yielded favorable outcomes, specifically a 97.7% accuracy rate for multi-class difficult backdrop data. The accuracy of the suggested Arabic SL system, employing the LBP descriptor, was only 9.78%, which may be compared to the findings presented by the authors [5].

both systems used images with identical requirements. The GLCM features yield the highest accuracy for photographs with intricate backgrounds, achieving a 90.2% classification rate using a KNN classifier and 10-fold cross-validation. In contrast, for images with black backgrounds, the accuracy is slightly lower at 88.3% with 5-fold cross-validation and a smaller dataset. The latter outcome implies that implementing any enhancement, such as transitioning to a 10-fold approach or expanding the dataset, would likely enhance accuracy. The system described in the reference [39] attained an accuracy of 85.3%, whereas the accuracy reported in the reference [5] was the lowest at 2.89% when using GLCM, indicating the presence of some type of error impact. Table 3 displays the three categories of Convolutional Neural Networks (CNNs) employed, along with the corresponding test outcomes for various data types. These findings indicate that SqueezeNet is the optimal choice due to its lower number of epochs and reduced storage requirements resulting from the compression of input, as demonstrated in Table 4.

The alteration in the quantity of epochs impacts the velocity and precision of outcomes in all varieties of CNNs. It is feasible to decrease the number of epochs based on photographs with simpler backgrounds. For SqueezeNet, a black background is utilized, necessitating only five epochs. The selection of CNN architecture also impacts the number of epochs necessary. To achieve satisfactory results with AlexNet, it was necessary to increase the number of epochs to 20 while dealing with photos that have complicated backgrounds. When using the same data, SqueezeNet achieved convergence in just 10 epochs. A comparative analysis was conducted in Barbhuiya *et al*., [11], to assess the performance of two convolutional neural networks (CNNs), namely AlexNet and VGG16. The accuracy of AlexNet was found to be 99.82% while using data sources with black backgrounds. This accuracy was slightly lower compared to the current usage of AlexNet, which achieved 100% accuracy. The authors employed GoogleNet in Sosa-Jimenez *et al.*, [42] to achieve a validation accuracy of almost 98% for five letters and 74% for 10 letters. The dataset they utilized included both black and complex backgrounds. The new work surpassed these results with an accuracy of 99.8% and 98.52% for black and complicated backgrounds, respectively. In the study referenced as Kasukurthi *et al*., [25], the researchers utilized SqueezeNet to achieve a maximum training accuracy of 87.47%. The corresponding validation accuracy reached 83.29%.

The corresponding outcomes in this instance are 100%, 99.85%, and 97.57% for black backgrounds, complex backgrounds, and the complete ASL alphabet, respectively. To achieve improved accuracy in handling various sorts of data, it is necessary to increase the number of photos for each class, particularly when dealing with complicated backgrounds. The accuracy achieved was 99.89% for a dataset size of 6×150, with a training time of 6.15 hours and a testing time of 11.2 seconds. In contrast, a dataset size of 27×50 resulted in an accuracy of 97.57%, with a training time of 17 hours and a testing time of 24.86 seconds. The decrease in the quantity of photos each gesture category, from 150 to 50, resulted in a decline in accuracy. Consequently, in order to enhance accuracy for the entire alphabet, it is imperative to raise this quantity to 150 images per gesture. However, this would extend training time well beyond the observed 17 hours, and the required storage space would also increase. Examples below of a different number of samples for (4-signs) can show this fact:

With 10 samples for every sign giving an accuracy reach of (46.6%), we can also notice fluctuations in the precision value in almost all periods as shown in Figure10.

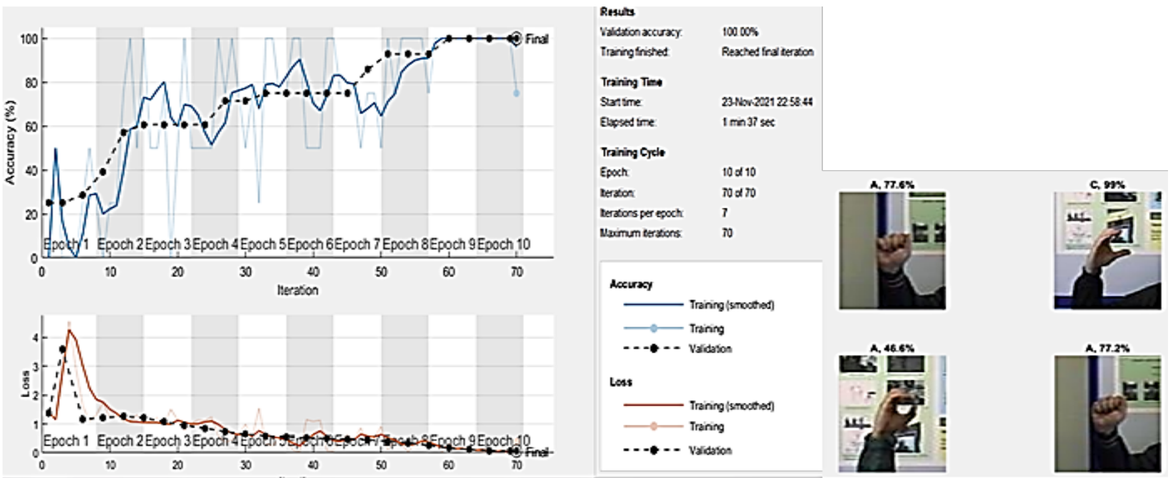

Fig. 10. The training progress of (4-signs) with 10 samples

Semarak Engineering Journal
Volume 11, Issue 1 (2025) 102-121

With 20, samples for every sign the accuracy value began to improve to the testing data as shown in Figure11.

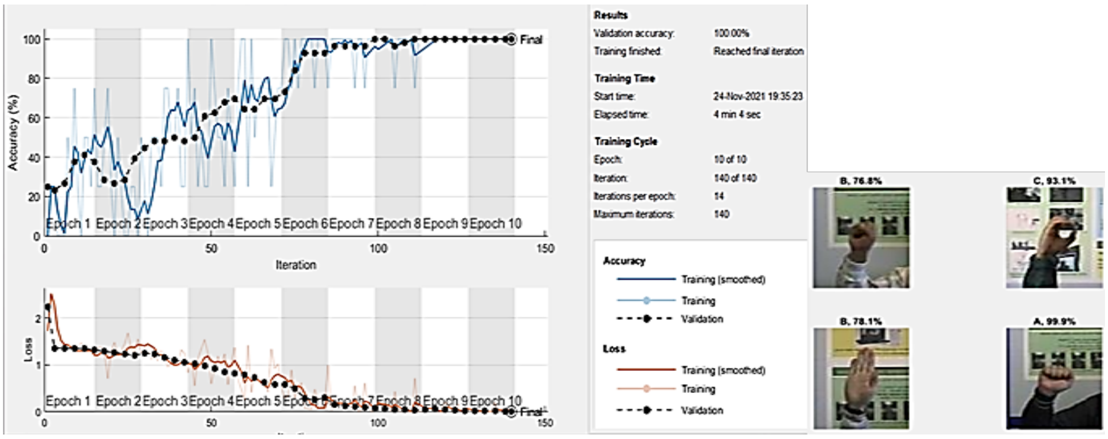

Fig. 11. The training progress of (4-signs) with 20 samples

With 40 sample for every sign the test, give a best result in term of accuracy but when repeating the test, there are some results that do not give high accuracy as shown in Figure 12.

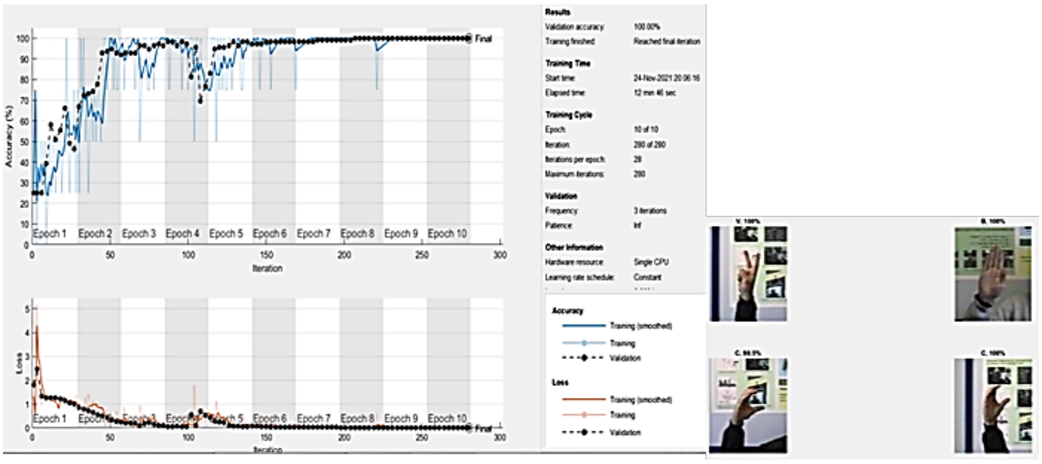

Fig.12. The training progress of (4-signs) with 40 samples

With 100 sample for every sign, the result was very suitable and still gave high accuracy whatever the test step is repeating, as shown in Figure 13. Which shows a smooth curve from the middle of the periods (5-epoch) traversed by the system.

Fig. 13. The training progress of (4-signs) with 100 samples

These test results can be demonstrated in terms of the number of samples for every class and the (Acc. %) for every image in the random choice tests with the showing of it is true (T) or false (F) in (Table 5).

Table 5
Comparing the results when changing the number of samples for each class

No samples	1st image	2nd image	3rd image	4th image
10	*77.6(F)*	*99(T)*	*46.6(F)*	*77.2(T)*
20	*76.8(F)*	*93.1(T)*	*78.1(T)*	*99.9(T)*
40	*100(T)*	*100(T)*	*98.5(T)*	*100(T)*
100	*100(T)*	*100(T)*	*100(T)*	*100(T)*

After testing the system in the off-time phase, HOG gave the best results but when the system was tested in real time the result was very pore so we chose the SqueezeNet model only, which offers the best results in the real-time phase. Therefore, the next step is utilizing the proposed system in the real-time phase for 6-singes, which will be used as orders in the smart environment. The results of recognition in terms of accuracy and the name of the class that the system selected appear on the upper side of the capturing image as an output.

The recognizer system is with a high level of accuracy in instantaneous time. The capturing images are taken in different conditions. The signs that chosen are represent the letters (A, B, C, five, V, and Point) in ASL, the images are with different distances from the camera and different persons. Some images are captured by rotating in direction and others with a lighting background. Although some images are with a complex background, which is not the same as the background for the dataset that the system learned on, the accuracy is still suitable and the recognition is right. These results and the average over all accuracy can be found from Table 6.

In this work, a commend voice, which represent the output text for the recognized image was included as an output for the system to be active in the smart environment system.

To compare with previous studies that used SqueezeNet, [25] it in off-time with the dataset called (surrey finger) and the validation accuracy was 83.29%.

In comparison with another system (R, N.M, and Abburi 2020), which converts gestures to a speech using the traditional machine learning TML system by extracting features using PCA and SVM, KNN classifiers to achieve 90% accuracy. The image of that system is limited to a white background and clear so the proposed system outperforms it due to the complex background for the images used.

A summary of SLR studies for static hand gestures by a vision-based approach in real-time with the utilized system either TML or Deep learning (DL) systems, background information, and the output type, is shown in Table 7.

Table 6
The average overall accuracy

Sign	Different forms of (Acc.%)					Average
A	*100*	*99.9*	*93.3*	*99.3*	*89.7*	*96.44*
B	*100*	*88.7*	*99.4*	*100*	*99.9*	*97.6*
C	*100*	*100*	*94*	*99.4*	*99.8*	*98.64*
V	*99.8*	*100*	*99.9*	*98.5*	*98.2*	*99.28*
Five	*100*	*100*	*99.7*	*99.3*	*100*	*99.8*
Point	*100*	*99.8*	*100*	*99.9*	*99.3*	*99.8*

Average over all accuracy = 98.593 %
With estimated time for every sign recognition = 0.001 sec

Table 7
Summary of SLR studies compared to the proposed system

Ref.	Dataset	year	preprocessing	FE	Classifier	Acc. %
			By TML systems			
(Dardas and Georganas 2011)	(10)-bare hand posture white wall background	2011	face subtract.& detect skin, hand gesture contour comparison	SIFT features via bag-of-features	K-means clustering SVM classifier	96.23 in 0.017 sec (text)
(Jayashree R. Pansare, Gawande, and Ingle 2012)	26-ASL alphabet In complex background	2012	detect hand skin convert to binary median filter Gaussian filter Morphological	edge detection by “Sobel” method	Euclidian distance	90 in 0.5 sec (text)
(Ahmad and Akhter 2013)	10- static hand gesture Complex background	2013	Subtract Background Color segmentation, Select(ROI)	Converted contour to orientation based on hash codes	hamming distance	82.1 in 7.36ms (text)
[34]	(BdSL)\2-hand Bengali language Uniform color background	2014	skin color segm. ,reduce noise by Gaussian conv. gray-BW	FE for hand shape :(finger position, fingertip)	KNN\ Vowels Consonants	(text) 98.17 & 94.75
(Jayshree R. Pansare and Ingle 2016)	ASL capturing by web camera Black background	2016	removing noise morphological operation Skin detector	(EOH) features	SVM	88.26 in 0.5sec (text)
			By DL systems			
(Islam, Siddiqua, and Afnan 2017)	ASL / Black background	2017	resized to 260×260 RGB to binary median filter rotated images	fingertip finder, eccentricity, elongatedness	ANN neural network	94.32 in 10sec (text)
(Taskiran, Killioglu, and Kahraman 2018)	ASL gestures Massey dataset Black background	2018	converted RGB to YCbCr convex hull algorithm for skin color	Hidden features extraction by CNN	Hidden classifier by CNN	98.05 (text)

(Bohra et al. 2019)	capturing 40 different gestures uniform color background	2019	Convert to grayscale Use median blur Hand detection by Color skin	maximum contoured area for hand gesture	CNN	99 in 0.0008sec (text)
(Kadhim and Khamees 2020)	(28-classes) ASL Different backgrounds	2020	resized (224 x 224)	Hidden features extraction by CNN:VGGnet	Hidden classifier VGGnet	98.65% (text)
(Mujahid et al. 2021)	gestures of five numbers: Uniform color background.	2021	UsingYOLO with200images increasing 2-fold, each image was duplicated	Obtaining boundary box (x-axis, y-axis, height, and width) in hidden layers: YOLOv3 & DarkNet-53	Hidden classifier in YOLOv3 and DarkNet-53 CNNnet	97.68% (text)
(Shin et al. 2023)	77 label KSL	2023		convolution and transformer-based multi-branch network	modified version of FFN	89.00%
Our proposed system	complex background Different hand shape (male &female& child)	2024	resized (227x227) only	No manual FE	Hidden classifier by SqeeseNet	98.5 in 0.001 sec

- ***Smart classroom application***

Modern civilization is undergoing a trend known as "smart environment technology" which creates intelligent living spaces for daily comfort and ease. Classrooms can be one of these environments that are designed to be automated structures supported by equipment for control, monitoring, and detection, including heating and cooling, lighting, ventilation, and security [38]. Gateways are the name given to these contemporary systems, which include sensors and switches and communicate via a central axis. These gateways are control systems that have a user interface on smartphones or computers. The Internet of Things (IoT) controls the communication network [2]. Smart classroom users can manage, control, and automate any electronic component in their environment because every digital gadget has connectivity to the Internet [50]. One of the voice-controlled applications is Amazon Alexa which was developed by the Amazon company. It is apparently used to play music, respond to general inquiries, set timers and alarms, or manage networked devices. Users submit a request, which Alexa filters using speech recognition and natural language processing. After gaining access to gateway services, Alexa responds to the user [28].

- **Smart classroom for deaf people by Alexa device**

An important question is when does Alexa start responding? In fact, Alexa is designed to listen continuously, but the response is only achieved when it hears a keyword to wake up. In the proposed system (Alexa) echo dot is used, which is provided by Amazon as a voice control device as shown in Figure 14. Six comments have been suggested for the 6-signs that have been identified here these comments are shown in (Table 8).

Table 8
Comments that chosen for the 6-signs

Sign	Comment
A	Alexa
B	open the window
C	turn on lamp
V	close the window
Five	turn on the air conditioner
Point	play alarm

Fig.14. (Alexa) echo dot from amazon

When the recognition process is done, Alexa needs these voice comments to achieve a response. So an algorithm for changing the text to voice is implemented by using the Speech Synthesizer net system the output is of a sound voice and text above the image gesture as shown in Figure15.

Fig. 15. Results with text commend

For implementing a smart environment service, there is a need for two application:
Amazon Alexa app., which is necessary to setup Alexa and control its services and skills. Smart life app. Which is necessary to specify the smart devices that the user want to be added to his smart environment control system. Alexa will employ the comment, to get certainty that Alexa is responding for the mute user there is a blue light ring will appear around the top of Alexa, as shown in Figure16. This response is surely not important only for deaf people but also for any user looking for more easy services for his smart environment.

Fig. 16. A blue light ring will spin around the top of Alexa at response

4. Conclusion

E-learning fosters individualized instruction, caters to various learning preferences, and streamlines communication, thus advancing equitable educational opportunities for kids who are Deaf-Mute.

This paper presents a fast and accurate system for sign recognition by focusing on higher accuracy, which would result in the use of a squeezenet CNN as proposed in this work. The study result revealed that this would cause the time for training to increase. Furthermore, the change in the number of epochs was identified as one of the major factors effecting the speed and accuracy of CNN results, irrespective of algorithm type. It was also noted that the reduction in the number of epochs depends on the properties of the image sample entered, such as image background complexity. In this study, a black background was used for SqueezeNet; there is a requirement for 5 epochs only, but when the image was with a complex background number of epochs was 10 epochs. The type of CNN used also has an effect. In AlexNet, the implementation of images with complex backgrounds required increasing the number of epochs to 20 to obtain satisfactory results. While in SqueezeNet with the same data, it only took 10 epochs. Based on these study findings, Squeezenet was revealed to be suitable for limited space memory because of its performance in squeeze element components, which gave an excellent idea for serving smart environment applications such as e-learning classrooms.

The suggested solution should be implemented in the future using a Raspberry Pi as the hardware component. Provide a reverse system that translates sound waves into Second Life visuals, more suited for usage in smart integrated environments that support deaf-dumped individuals in need of specialized, targeted healthcare.

Acknowledgement
This research was not funded by any grant.

References
[1] Abdul Malek Saidina Omar, Ahmad Asri Abd Samat, Fasihah Faisal, Muhammad Khusairi Osman, Mohammad Nizam Ibrahim, and Zakaria Hussain. 2023. “Convolutional Neural Network for Transmission Line Fault Diagnosis Based on Signal Segmentation Approach.” *Journal of Advanced Research in Applied Sciences and Engineering Technology* 32 (3): 199–216. https://doi.org/10.37934/araset.32.3.199216

[2] Adam Wong Yoon Khang, Jamil Abedalrahim Jamil Alsayaydeh, Johar Akbar Mohamat Gani, Tarani Bascar, Azman Awang Teh, Jaysuman Pusppanathan, and Andrii Oliinyk. 2024. “Portable American Sign Language System Using RF Signals and IoT Technology for Deaf-Blind People.” *Journal of Advanced Research in Applied Sciences and Engineering Technology* 45 (1): 79–89. https://doi.org/10.37934/araset.45.1.7989

[3] Adeyanju, Ibrahim Adepoju, Oluwaseyi Olawale Bello, and Mutiu Adesina Adegboye. "Machine learning methods for sign language recognition: A critical review and analysis." *Intelligent Systems with Applications* 12 (2021): 200056. https://doi.org/10.1016/j.iswa.2021.200056

[4] Ahmad, Saleh Ud-din, and Shamim Akhter. "Real time rotation invariant static hand gesture recognition using an orientation based hash code." In *2013 International Conference on Informatics, Electronics and Vision (ICIEV)*, pp. 1-6. IEEE, 2013. https://doi.org/10.1109/ICIEV.2013.6572620

[5] Alzohairi, Reema, Raghad Alghonaim, Waad Alshehri, and Shahad Aloqeely. "Image based Arabic sign language recognition system." *International Journal of Advanced Computer Science and Applications* 9, no. 3 (2018). https://doi.org/10.14569/IJACSA.2018.090327

[6] “American Sign Language Detection Using PCA and LDA - File Exchange - MATLAB Central.” n.d. https://uk.mathworks.com/matlabcentral/fileexchange/36394-american-sign-language-detection-using-pca-and-lda.

[7] Munanday, Anbananthan Pillai, Norazlianie Sazali, Arjun Asogan, Devarajan Ramasamy, and Ahmad Shahir Jamaludin. "The implementation of transfer learning by convolution neural network (CNN) for recognizing facial emotions." *Journal of Advanced Research in Applied Sciences and Engineering Technology* 32, no. 2 (2023): 255-

276. https://doi.org/10.37934/araset.32.2.255276
[8] Andi Muhammad Amil Siddik, Ainun Mawaddah Abdal, Armin Lawi, and Edy Saputra Rusdi. 2024. "Ensemble Transfer Learning for Hand-Sign Digit Image Classification." *Journal of Advanced Research in Applied Sciences and Engineering Technology* 43 (1): 95–111. https://doi.org/10.37934/araset.43.1.95111
[9] Ansar, Hira, Ahmad Jalal, Munkhjargal Gochoo, and Kibum Kim. "Hand gesture recognition based on auto-landmark localization and reweighted genetic algorithm for healthcare muscle activities." *Sustainability* 13, no. 5 (2021): 2961. https://doi.org/10.3390/su13052961
[10] "ASL Alphabet | Kaggle." n.d. Accessed September 20, 2022. https://www.kaggle.com/datasets/grassknoted/asl-alphabet?select=asl_alphabet_train.
[11] Barbhuiya, Abul Abbas, Ram Kumar Karsh, and Rahul Jain. "CNN based feature extraction and classification for sign language." *Multimedia Tools and Applications* 80, no. 2 (2021): 3051-3069. https://doi.org/10.1007/s11042-020-09829-y. https://doi.org/10.1007/s11042-020-09829-y
[12] Barczak, Andre LC, Napoleon H. Reyes, Maria Abastillas, Ana Piccio, and Teo Susnjak. "A new 2D static hand gesture colour image dataset for ASL gestures." (2011).
[13] Bheda, Vivek, and Dianna Radpour. "Using deep convolutional networks for gesture recognition in american sign language." *arXiv preprint arXiv:1710.06836* (2017).
[14] Bodda, Sai Charan, Palki Gupta, Gaurav Joshi, and Ayush Chaturvedi. "A new architecture for hand-worn Sign language to Speech translator." *arXiv preprint arXiv:2009.03988* (2020).
[15] Bohra, Tanuj, Shaunak Sompura, Krish Parekh, and Purva Raut. "Real-time two way communication system for speech and hearing impaired using computer vision and deep learning." In *2019 International Conference on Smart Systems and Inventive Technology (ICSSIT)*, pp. 734-739. IEEE, 2019. https://doi.org/10.1109/ICSSIT46314.2019.8987908
[16] Dardas, Nasser H, and Nicolas D Georganas. 2011. "Real-Time Hand Gesture Detection and Recognition.Pdf" 60 (11): 3592–3607. https://doi.org/10.1109/TIM.2011.2161140
[17] Ekbote, Juhi, and Mahasweta Joshi. 2016. "A Survey on Hand Gesture Recognition for Indian Sign Language," 1039–44.
[18] Haque, Promila, Badhon Das, and Nazmun Nahar Kaspy. "Two-handed bangla sign language recognition using principal component analysis (PCA) and KNN algorithm." In *2019 international conference on Electrical, Computer and Communication Engineering (ECCE)*, pp. 1-4. IEEE, 2019. https://doi.org/10.1109/ECACE.2019.8679185
[19] Hassanpour, M., and H. Malek. "Learning document image features with SqueezeNet convolutional neural network." (2020): 1201-1207. https://doi.org/10.5829/ije.2020.33.07a.05
[20] Hussain, Zawar, Quan Z. Sheng, and Wei Emma Zhang. "A review and categorization of techniques on device-free human activity recognition." *Journal of Network and Computer Applications* 167 (2020): 102738. https://doi.org/10.1016/j.jnca.2020.102738
[21] Islam, Md Mohiminul, Sarah Siddiqua, and Jawata Afnan. "Real time hand gesture recognition using different algorithms based on American sign language." In *2017 IEEE international conference on imaging, vision & pattern recognition (icIVPR)*, pp. 1-6. IEEE, 2017. https://doi.org/10.1109/ICIVPR.2017.7890854
[22] Jaramillo-Yánez, Andrés, Marco E. Benalcázar, and Elisa Mena-Maldonado. "Real-time hand gesture recognition using surface electromyography and machine learning: A systematic literature review." *Sensors* 20, no. 9 (2020): 2467. https://doi.org/10.3390/s20092467
[23] Jiang, Xianwei. "Isolated Chinese sign language recognition using gray-level Co-occurrence Matrix and parameter-optimized Medium Gaussian support vector machine." In *Frontiers in Intelligent Computing: Theory and Applications: Proceedings of the 7th International Conference on FICTA (2018), Volume 2*, pp. 182-193. Singapore: Springer Singapore, 2019. https://doi.org/10.1007/978-981-13-9920-6_19
[24] Kadhim, Rasha Amer, and Muntadher Khamees. "A real-time american sign language recognition system using convolutional neural network for real datasets." *Tem Journal* 9, no. 3 (2020): 937. https://doi.org/10.18421/TEM93-14
[25] Kasukurthi, Nikhil, Brij Rokad, Shiv Bidani, and Aju Dennisan. "American sign language alphabet recognition using deep learning." *arXiv preprint arXiv:1905.05487* (2019).
[26] Kim, Jungi, and Patricia O'Neill-Brown. "Improving American sign language recognition with synthetic data." In *Proceedings of Machine Translation Summit XVII: Research Track*, pp. 151-161. 2019.
[27] Kraljević, Luka, Mladen Russo, Matija Pauković, and Matko Šarić. "A dynamic gesture recognition interface for smart home control based on croatian sign language." *Applied Sciences* 10, no. 7 (2020): 2300. https://doi.org/10.3390/app10072300
[28] Lopatovska, Irene, Katrina Rink, Ian Knight, Kieran Raines, Kevin Cosenza, Harriet Williams, Perachya Sorsche, David Hirsch, Qi Li, and Adrianna Martinez. "Talk to me: Exploring user interactions with the Amazon Alexa." *Journal of Librarianship and Information Science* 51, no. 4 (2019): 984-997. https://doi.org/10.1177/0961000618759414

[29] Masud, Mohd Akmal, Mohd Zamani Ngali, Siti Amira Othman, Ishkrizat Taib, Kahar Osman, Salihatun Md Salleh, Ahmad Zahran Md Khudzari, and Nor Salita Ali. "Variation Segmentation Layer in Deep Learning Network for SPECT Images Lesion Segmentation." *J. Adv. Res. Appl. Sci. Eng. Technol* 36 (2024): 83-92. https://doi.org/10.37934/araset.36.1.8392

[30] Molchanov, Pavlo, Xiaodong Yang, Shalini Gupta, Kihwan Kim, Stephen Tyree, and Jan Kautz. "Online detection and classification of dynamic hand gestures with recurrent 3d convolutional neural network." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4207-4215. 2016. https://doi.org/10.1109/CVPR.2016.456

[31] Mujahid, Abdullah, Mazhar Javed Awan, Awais Yasin, Mazin Abed Mohammed, Robertas Damaševičius, Rytis Maskeliūnas, and Karrar Hameed Abdulkareem. "Real-time hand gesture recognition based on deep learning YOLOv3 model." *Applied Sciences* 11, no. 9 (2021): 4164. https://doi.org/10.3390/app11094164

[32] Duodu, Nana Yaw, Warish D. Patel, and Hakan Koyuncu. "Advancements in Telehealth: Enhancing breast cancer detection and health automation through smart integration of iot and cnn deep learning in residential and healthcare settings." *J. Adv. Res. Appl. Sci. Eng. Technol* 45, no. 2 (2025): 214-226. https://doi.org/10.37934/araset.45.2.214226

[33] Pansare, Jayashree R., Shravan H. Gawande, and Maya Ingle. "Real-time static hand gesture recognition for American Sign Language (ASL) in complex background." *Journal of Signal and Information Processing* 3, no. 3 (2012): 364-367. https://doi.org/10.4236/jsip.2012.33047

[34] Pansare, Jayshree R., and Maya Ingle. "Vision-based approach for American sign language recognition using edge orientation histogram." In *2016 international conference on image, vision and computing (ICIVC)*, pp. 86-90. IEEE, 2016. https://doi.org/10.1109/ICIVC.2016.7571278

[35] R, Priyakanth, Sai Krishna N.M, and Radha Abburi. 2020. "Hand Gesture Recognition and Voice Conversion for Speech Impaired." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3734777

[36] Raghuveera, Tripuraribhatla, R. Deepthi, R. Mangalashri, and R. Akshaya. "A depth-based Indian sign language recognition using microsoft kinect." *Sādhanā* 45, no. 1 (2020): 34. https://doi.org/10.1007/s12046-019-1250-6

[37] Rahaman, Muhammad Aminur, Mahmood Jasim, Md Haider Ali, and Md Hasanuzzaman. "Real-time computer vision-based Bengali sign language recognition." In *2014 17th international conference on computer and information technology (ICCIT)*, pp. 192-197. IEEE, 2014. https://doi.org/10.1109/ICCITechn.2014.7073150

[38] Mahariya, Satish Kumar, Awaneesh Kumar, Rajesh Singh, Anita Gehlot, Shaik Vaseem Akram, Bhekisipho Twala, Mohammed Ismail Iqbal, and Neeraj Priyadarshi. "Smart campus 4.0: Digitalization of university campus with assimilation of industry 4.0 for innovation and sustainability." *J. Adv. Res. Appl. Sci. Eng. Technol* 32, no. 1 (2023): 120-138. https://doi.org/10.37934/araset.32.1.120138

[39] Savur, Celal, and Ferat Sahin. "American Sign Language Recognition system by using surface EMG signal." In *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 002872-002877. IEEE, 2016. https://doi.org/10.1109/SMC.2016.7844675

[40] Mirza, Muhammad Shaheer, Sheikh Muhammad Munaf, Fahad Azim, Shahid Ali, and Saad Jawaid Khan. "Vision-based Pakistani sign language recognition using bag-of-words and support vector machines." *Scientific Reports* 12, no. 1 (2022): 21325. https://doi.org/10.21203/rs.3.rs-1204236/v1

[41] Shin, Jungpil, Abu Saleh Musa Miah, Md Al Mehedi Hasan, Koki Hirooka, Kota Suzuki, Hyoun-Sup Lee, and Si-Woong Jang. "Korean sign language recognition using transformer-based deep neural network." *Applied Sciences* 13, no. 5 (2023): 3029. https://doi.org/10.3390/app13053029

[42] Sosa-Jiménez, Candy Obdulia, Homero Vladimir Ríos-Figueroa, Ericka Janet Rechy-Ramírez, Antonio Marin-Hernandez, and Ana Luisa Solís González-Cosío. "Real-time mexican sign language recognition." In *2017 IEEE international autumn meeting on power, electronics and computing (ROPEC)*, pp. 1-6. IEEE, 2017. https://doi.org/10.1109/ROPEC.2017.8261606

[43] S Suleiman, Ahmad A. "Image Classification Based on Enhancement of Local Binary Pattern." PhD diss., Middle East University, 2018.

[44] Taskiran, Murat, Mehmet Killioglu, and Nihan Kahraman. "A real-time system for recognition of American sign language by using deep learning." In *2018 41st international conference on telecommunications and signal processing (TSP)*, pp. 1-5. IEEE, 2018. https://doi.org/10.1109/TSP.2018.8441304

[45] The Mathworks. 2018. "Introducing Deep Learning with MATLAB: 'What Is Deep Learning?'" *The MathWorks, Inc. MATLAB and Simulink Are Registered Trademarks of The MathWorks, Inc*, 1–14.

[46] Ullah, Mohib, Ahmed Mohammed, and Faouzi Alaya Cheikh. "Pednet: A spatio-temporal deep convolutional neural network for pedestrian segmentation." *Journal of Imaging* 4, no. 9 (2018): 107. https://doi.org/10.3390/jimaging4090107

[47] Vinoth, N., and K. Nirmala. "Deaf students higher education system using e-learning." *Journal of Education and Learning (EduLearn)* 11, no. 1 (2017): 41-46. https://doi.org/10.11591/edulearn.v11i1.5131

[48] Wang, Aili, Minhui Wang, Kaiyuan Jiang, Mengqing Cao, and Yuji Iwahori. "A dual neural architecture combined SqueezeNet with OctConv for LiDAR data classification." *Sensors* 19, no. 22 (2019): 4927. https://doi.org/10.3390/s19224927

[49] Ashani, Zahra Nazemi, Iszuanie Syafidza Che Ilias, Keng Yap Ng, Muhammad Rezal Kamel Ariffin, Ahmad Dahari Jarno, and Nor Zarina Zamri. "Comparative analysis of deepfake image detection method using vgg16 vgg19 and resnet50." *Journal of Advanced Research in Applied Sciences and Engineering Technology* 47, no. 1 (2024): 16-28. https://doi.org/10.37934/araset.47.1.1628

[50] Z Zaidan, A. A., and B. B. Zaidan. "A review on intelligent process for smart home applications based on IoT: coherent taxonomy, motivation, open challenges, and recommendations." *Artificial Intelligence Review* 53, no. 1 (2020): 141-165. https://doi.org/10.1007/s10462-018-9648-9