PLOMS Journal of Artificial Intelligence (PLOMS AI)

PLOMS AI

Real-time Arabic sign language translator Using media pipe and LSTM

Sahar K. Hussin[1]; Omar Mohamed[2]; Mustafa Mohamed[2]; Eslam Ahmed [2]; Omar Mahmoud [2]
[1,2]Communication and Computers Engineering Department Alshorouck Academy, Cairo, Egypt
Correspondence :s.kamal@sha.edu.eg

Abstract: People who are dumb and deaf have difficulty communicating daily. Artificial intelligence (AI) developments have allowed the breaking down of this communication barrier. As a result of this work, an Arabic sign language (ASLT) letter recognition system has been created. The ASLT recognition system employs BILSTM with a media pipe structure to interpret depth data and enhance the social interaction of hearing-impaired people. Depending on user input, the suggested approach would automatically detect and identify Arabic and hand-sign alphabet letters. The proposed model should have a 98% accuracy rate in identifying ASLT for letters, 96% for words, and 100% for digits from 0 to 9. We conducted comparative research to evaluate our method, and the results showed that it is more accurate at differentiating between static signs than earlier studies using the same dataset.

Keywords: sign language translator, deep learning, BiLSTM media pipe.

I. Introduction

Using facial and bodily expressions, postures, and gestures in sign language allows for human-to-human conversation and communication on TV and social media. Millions of deaf (hearing impaired) persons and those who are hard of hearing and have a variety of speech challenges use sign language as their first language. The British Deaf Association's investigation revealed that around 151,000 persons use sign language to communicate [1]. Since practically every nation has its own sign language and finger spelling alphabet, there is no universal sign language. They replicate facial expressions and animate their lips in addition to using manual gestures. Around 22 Arab nations use various gestures in Arabic Sign Language (ArSL). The cultural variance between these nations can be used as an explanation for the discrepancy observed for specific word motions. The motions for the Arabic letters and numerals are the same throughout all 22 nations, notwithstanding the absence of standardization in ArSL [3]. The Arabic alphabets are expressed using the standard Arabic sign language, as seen in Fig. 1.

The visual form of sign language allows for quick and precise information exchange between individuals. It's crucial to spell words correctly and accurately translate ideas and emotions in a brief time. Creating a vision-based sign language translator is crucial since some people struggle to understand sign language, and some do not. Creating such a system makes it possible to lower the communication gap between individuals drastically. There are two main ways to translate sign language. The first is vision-based techniques, which use the onboard camera(s) to take the desired pictures, which are then sent to the image analysis module [4-7]. The second strategy is a glove-based strategy that implements sensors and gloves. In this case, the additional hardware (glove) is used to overcome the limitations of the conventional vision-based approaches. Glove-based approaches often feel burdensome and intrusive to signers and users, but they produce outcomes that are more accurate and exact [8]. In this study, we offer a vision-based sign language translation technique that dynamically records hand motions using a single video camera. There are two main parts to the Arabic sign language. The first part is a complete language where each word is represented by a sign (for instance, the word father is represented by a sign). Arabic sign languages vary among Arab nations; examples are Saudi and

Egyptian sign languages. The first section is referred to as Arabic Sign Language (ArSL). The Arabic sign language alphabet represents each letter of the Arabic alphabet in the second section (ArSLA). Due to the significance of Arabic sign language, which was previously discussed, researchers and practitioners have become interested in the problem of creating Arabic sign language recognition systems. The literature offers numerous answers for both ArSL and ArSLA. Generally, sign language systems have five major stages, with each level serving a specific purpose. These five steps are image acquisition, image preprocessing, feature extraction from those images, image segmentation, and classification process. Numerous methods have been presented to identify sign language gestures. To address this issue, transfer learning [4], a method in which the model is trained on a large training set, has been suggested. Afterward, the training's outcomes are used as the task's beginning point. In disciplines like language processing and computer vision, transfer learning is effective. Data augmentation is another method discovered to reduce over-fitting and enhance performance [5]. The training set's size is increased using this technique by applying geometric and color adjustments, such as rotation, resizing, cropping, adding noise to the image, blurring it, etc.

This study uses media pipe to demonstrate a real-time recognition model for the Arabic Sign Language Alphabet (ASLT)). The major goal was to create a solution that could be used by everyone.

Our contributions are summarized as follows:

1. Our aim in building a machine translation system from Arabic sign Language is to facilitate.
   the communication between hearing people and deaf people. Deaf people can use the system as a chat tool.
2. We developed an efficient and powerful real-time Arabic Sign Language letter, Word, and number recognition model using media pipe and BiLSTM.
3. It allows everyone to use it through online group video calls.
4. The proposed model was developed with the complete Arabic alphabet, which allows users to write full articles using Arabic Sign Language letters in real time.
5. This model's great challenge is detecting more than one object (a hand) without considering the background. The background of the hand plays a prominent role in object recognition.

II. Related Work

This section studies several approaches that have been presented for sign language gesture identification. According to [10], sign language recognition systems can be categorized into glove-based and deep learning-based systems. The first category is based on hardware devices that are made up of unique sensors that can be packaged in various forms that are suited for manual usage (since sign languages are characterized by hands). The second group relies solely on the camera. Despite the first group's promising results, the second group is still the better option because all it needs in terms of technology is a camera, which can be found in practically any modern computer. The first category is referred to in the literature as sensors-based solutions, and the second is image-based solutions. In this paper, we follow the concepts of the second group.

Al-Nuaimy [11] proposed a low-cost electronic glove to communicate between deaf and normal people. According to the authors, a force flex sensor in the glove would translate a hand press into an audible and visible signal. The system comprised three Flex Force sensors, an Arduino Uno microprocessor, a display, and a speaker. Nevertheless, it was tested only for six words. Lee et al. [12] created a smart hand-wearable sign gesture interpreter to recognize American Sign Language (ASL) characters.

The system comprised three components: a processor unit, a unit for mobile applications, and a unit for hand-worn sensors. Results of the experiment showed that the system's accuracy was 65.7% without pressure sensors and improved to 98.2% when a pressure sensor was employed on the middle finger. Kala et al. created and deployed a hand gesture translator embedded device [13] to bridge the communication gap between mute and

non-mute people. Three components comprise the developed model: a sensor component, input processing, and output communication. The system included an Android smartphone, an LCD, an Arduino ATMEGA 2560, a three-axis accelerometer, and flex sensors mounted to each glove finger. However, the device is not portable, and it is powered by the mains, so a voice translator for the Indian Sign language was created. This device uses five flex sensors and an inertia measurement unit to record the gesture (IMU).

In the second group, many papers proposed using deep learning models to translate sign language. Kamruzzaman proposed an ArSLA recognition system to translate signs into Arabic speech [14]. Convolutional neural networks were used to create the classifier (CNN).

The research in [15] created a comparison survey to examine how well ArSLA classifiers function. CNN, RNN, MLP, LDA, HMM, ANN, SVM, and KNN were the classifiers used. An alternative ArSL recognition system built on deep CNN was shown in [16]. The system aims to minimize the number of parameters to extract and detect hand movements—a collection of 50,000 photos with signs made by a population of signers of different ages. The system's top result in their testing was a 97% accuracy rate. This paper compares our model with three papers with the same data set. In [17], we offer a lightweight Efficient Network (Efficient Net)-based image-based ArSL recognition system to improve overall ArSL identification performance while decreasing the number of parameters and time complexity. The proposed system's performance will be assessed using accepted performance metrics and actual datasets in their testing; the system's top result was a 94.3% accuracy rate. Unprecedented research efforts have been conducted in Arabic Sign Language to identify hand motions and signals using a deep learning model. A vision-based system to recognize Arabic characters based on hand movements and convert them into Arabic speech is proposed by [23]. Using a deep learning model, the suggested system will automatically recognize letters made of hand signs and pronounce the result in Arabic. This technique offers up to 90% accuracy. Although [23] mentions the digit model, it doesn't show the accuracy of digit translation. To translate Arabic (ArSL), [24] suggests ArSLCNN, a deep learning model based on a convolutional neural network (CNN). The ArSL2018 dataset, which contains 54,049 photos of 32 sign language motions taken from 40 people, was used in experiments. The train and test accuracy in the first experiments with the ArSL-CNN model was 98.80% and 96.59%, respectively. The dataset was subjected to various re-sampling techniques for the second set of studies. The application of the synthetic minority oversampling technique (SMOTE) increased the total test accuracy from 96.59% to 97.29%, according to the results. Arabic and hand-signed alphabets will be automatically detected and identified in the suggested model [25], depending on user input. The proposed model should have a 97.1% accuracy rate in identifying ArSL.

III. **Proposed method**

The proposed system for the ASLT for letters and numbers is shown in Figure 1. The proposed system recognizes the Arabic sign language of alphabets and numbers
It recognizes gestures with motion and gestures without motion as well. The first step is to collect the dataset for the system, which will be discussed in the results and discussion section. Media Pipe Hands is a high-fidelity hand and finger tracking framework used in dataset preparation. It makes use of several ML models to infer the 21 three-dimensional landmarks of a hand in real time from just one frame. This mode consists of three stages. The proposed system consists of three stages: hand tracking, pre-processing, and training & testing. First, Media Pipe's pre-trained hand-tracking model extracts hand landmarks from sign language images or videos. Then, features are extracted from the hand landmarks, including hand shape, hand movement, and hand orientation. Finally, predication is done using different machine learning Algorithms

A. Stage 1: Pre-Processing of Images to Get Multi-hand Landmarks

The media pipe [21] solution has an ML pipeline at its back end and consists of two independent models: a) a palm detection model and b) a Landmark model. The Palm Detection model provides a precisely cropped image of the palm that is then transferred to the historical model. This process reduces data augmentation (such as rotation, flipping, and scaling) performed in deep learning models and devotes most of its power to feature localization. The traditional method detects the hand from the frame and then localizes the features to the current frame. However, in this Palm Detector, the ML pipeline uses challenges with a different strategy.

After the palm detection screen is scanned over the entire image frame, subsequent Hand Landmark models appear in the image. This model accurately locates 21 3D hand articulatory coordinates (i.e., x, y, z axis) within the detected hand regions. The model is so well-trained and powerful in hand detection that it coordinates with the partially visible hand. Figure 2 shows the Hand Landmark model's detection of 21 salient points.

Now that we have a functional model for palm and hand detection, this model is passed through our dataset with Arabic. We have the Alif to Ya alphabet in the Arabic Sign Language dataset. Therefore, we pass our detection model over each alphabetical folder containing images and perform manual detection, which yields 21 landmarks, as shown in Figure 2. The obtained feature points are then stored in a CSV file. A synchronous delete task is executed while feature points are being extracted. Here, only the x and y coordinates detected by the Hand Landmark model are considered for training the ML model. Depending on the size of the data set, about 10-15 minutes are required for Landmark extraction.

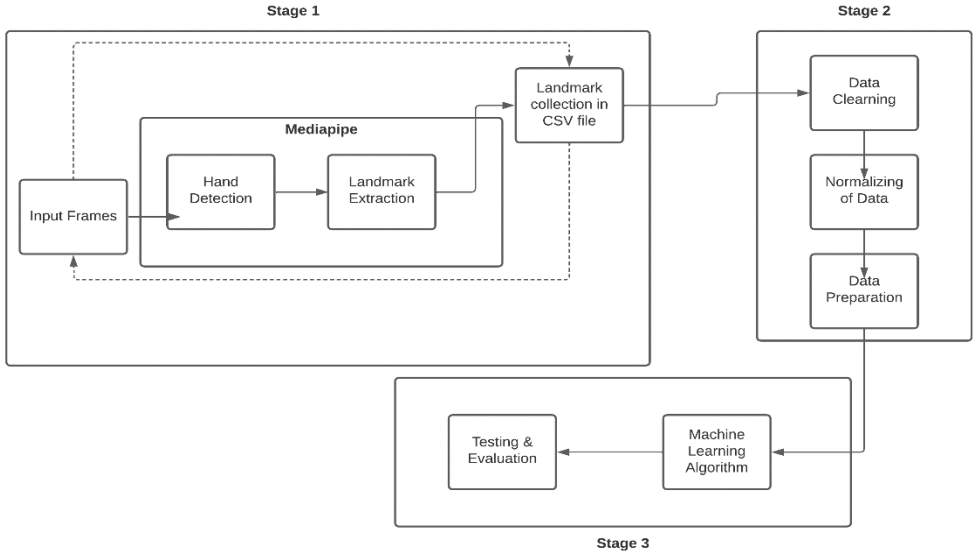


Figure 1. The proposed architecture of Arabic sign letter translation.

B. Stage 2: Data cleaning and normalization

Each image in the dataset is processed through stage 1 to collect all the data points under one file; just as in stage 1, we are just examining x and y coordinates from the detector. The panda's library function is then used to scrape this file and look for any null entries. Sometimes, a blurry image prevents the detector from seeing the hand, resulting in a null entry into the dataset. As a result, these issues must be cleared up; otherwise, bias will be introduced into the predictive model. Utilizing their indexes, rows with these null entries are looked for and deleted from the table. We normalized the x and y coordinates after eliminating spots that wouldn't fit our scheme. After that, the data file is ready to be divided into training and validation sets. 20% of the data is held back for model validation, while 80% is used to train our model using different optimization and loss functions.

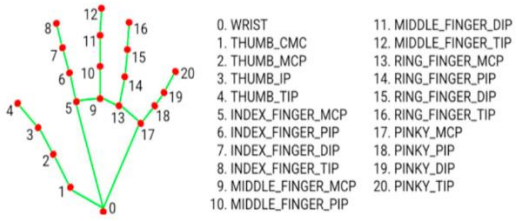


Figure 2. Hand landmarks representation used in media pipe [26].

C. Stage 3: Prediction using ML Algorithm

This paper used different machine Learning algorithms) to recognize Arabic sign language letters and numbers. Predictive analysis for different sign languages is performed using different machine learning algorithms such as artificial neural network (ANN), Support Vector Machine (SVM), K- nearest neighbor (KNN), Random Forest (RF), and decision tree (DT). Details of the analysis are discussed in Table 2 in the results section. ANN is effective in high-dimensional spaces. ANN works effectively when the number of samples is greater than

the number of dimensions. ANN is a set of supervised learning methods capable of classifying, regression, and detecting outliers. In this paper, we utilized Media pipe for hand key point capturing and Tensor Flow to train and detect the machine learning algorithm. The media pipe can also operate on both CPUs and GPUs; no extra processing power is required.

IV. Proposed model 2

The proposed model2 for the ASLT for traditional words is shown in Figure 5. The proposed model2 consists of three stages, stage1-1: Hand Landmarks using Media pipe, same as the previous model. Stage 1-2 Pose Landmarks using Media pipe. Each image in the dataset is sent through stage 1 to aggregate all the data points under one array since in stages 1-1 and 1-2, we are just considering the x, y, and z coordinates from the detector. Build sequences of landmarks for each word; each word consists of 40 frames, and each frame has hand and pose landmarks. Stage 2: prepossessing, which is the output of two previous stages containing all the sequences in one file. This file is then scraped through the pandas' library function to check for any null entries. After removing unwanted points, we normalized x, y, and z coordinates to fit into our system. The data file is then prepared for splitting into training and validation sets. Stage 3: Predict using Bidirectional LSTM, which is particularly suitable for recognizing continuous hand motion in ASLT hand gestures, as opposed to single static poses. Bidirectional LSTM (BiLSTM) is a type of LSTM that processes the input sequence in both forward and backward directions. BiLSTM can capture dependencies separated by long time intervals in both directions, making it more effective at modeling sequential data.

(a) **(b)**

Figure 3.**:** (a) pose landmarks representation used in media pipe [26]; (b) Bidirectional LSTM.

our model is designed to classify 10 words. Each word has 50 videos, and each video has 40 frames called Sequence. The model consists of two Bidirectional LSTM layers and a Dense layer to the Sequential model.

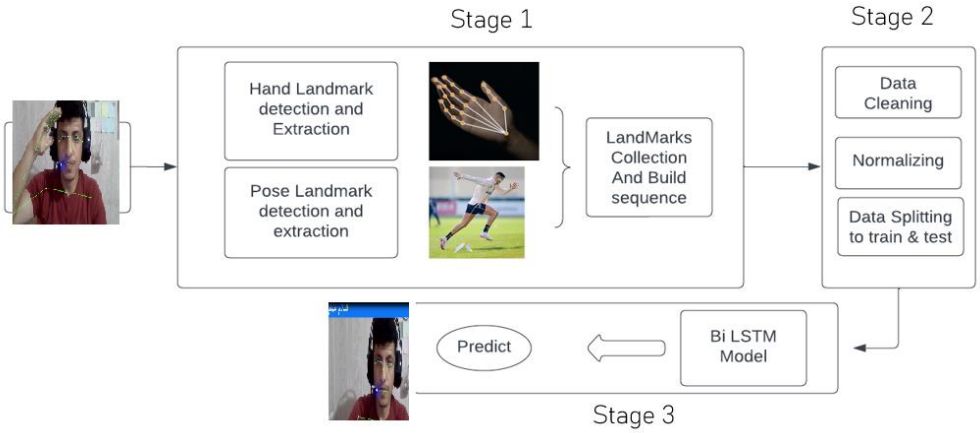


Figure 4. The proposed architecture of Arabic Sign Words translation.

V. Dataset

The dataset required for the experimental study of the proposed algorithm is created using Media Pipe, which is an open-source library. It captures the hand gestures for 40 frames and stores them as a Numpy array for training the model. The dataset has 1000 different permutations of a single alphabet. This makes a total of 32,000 numpy arrays; these are indexed and stored. The dataset has different parameters: the x, y, and z axes of the hand joints. The dataset is divided into 80% for training and 20% for testing with 10-fold cross-validation. Figure 3 shows samples of images, signs, and letters dataset.

Figure 5. Gesture dataset samples ASL

VI. MODEL TRAINING AND RESULTS

A. *Performance*

The proposal used Keras libraries and Python programming language that runs on the TensorFlow backend. Our model was trained on a machine that has an NVIDIA K80 graphics processing unit (GPU), 8 GB random access memory, and a 500 GB solid-state drive. The accuracy metric was adopted to determine the efficiency of our proposal. In formula (1), A denotes the accuracy, Tp and TN represent the number of correctly and incorrectly classified instances, respectively. The calculated value is multiplied by 100 to turn it into a percentage. To examine the performance of the proposed models, four useful metrics (Accuracy, Precision, Recall, and F1Score) are selected as follows:

A true positive result is one where the model accurately identifies the positive class. Similar to a true positive, a true negative is a result where the model accurately predicted the negative class. A false positive is an outcome where the model forecasts the positive class inaccurately. A false negative results when the model wrongly predicts the negative class. Accuracy is described as follows.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Where the ratio of True Positives to All Positives is known as precision.

$$precision = (TP)/\ (TP) + (FP) \quad (2)$$

The recall is the measure of the model correctly identifying True Positives.

$$Recall = (TP)/\ (TP) + (FN) \quad (3)$$

The F-score is calculated from the precision and recall of the test. The F-score or F-measure is a measure of a test's accuracy. The number of actual occurrences of the class in the specified dataset is Support.

B. *Results of Letter Recognition and Numbers*

In this discussion, the recognition model based on media pipe architecture has been developed and is ready to be tested and used. Now, Figures 4 show Arabic letters have been captured in real time. A gestures technique

has been used for identifying and capturing hand shape. In image processing, a gesture is defined as a technique in which part of the human body is recognized by using a camera.

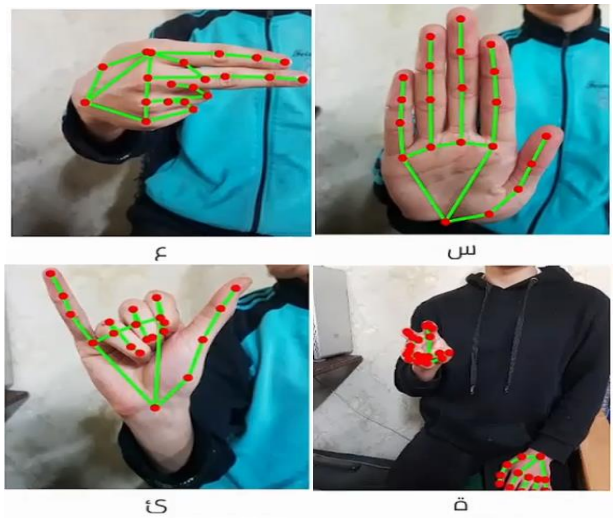


Figure 6. Real-time Arabic Sign Letters Recognition

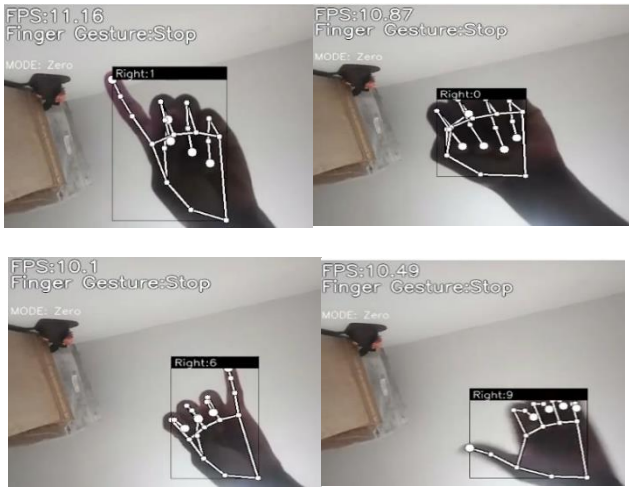


Figure 7. Real-time Arabic Number Sign Language Recognition

Table I indicates the accuracy of all 32 classes. From the table it can be observed that the number of testing samples across the classes varies considerably. It can also be observed that all classes have the same number so all classes have accuracy related to each other except 'ه' and 'ح' class contains 113 testing samples, and its accuracy was 91% and 93% respectively, these results revealed that quality of image and shape of letter and the imbalanced distribution of the number of samples between classes. The highest accuracy of about 100% was obtained for the letter(ح), as it is shown in Table 3 and figure 6 However, the lowest accuracy of about 91% was obtained for the letter (ه).

Table 1. Testing accuracy for each label of 32-classes

Letter	Precision	Recall	F1-score	Support
أ	0.97	0.99	0.98	91
ب	0.99	1.00	1.00	117
ت	0.94	0.98	0.96	103
ث	0.97	1.00	0.98	92
ج	0.93	0.97	0.95	102
ح	1	1	1	113
خ	0.98	0.97	0.98	117

د	0.99	0.92	0.95	88
ذ	0.97	1.00	0.98	91
ر	0.96	0.98	0.97	97
ز	1.00	0.98	0.99	112
س	0.97	0.98	0.97	114
ش	0.99	0.95	0.97	93
ص	1.00	0.99	1	124
ض	0.97	0.99	0.98	94
ط	0.99	0.99	0.99	119
ظ	0.98	0.92	0.95	91
ع	0.95	0.96	0.95	99
غ	1.00	0.98	0.99	105
ف	0.98	0.97	0.98	110
ق	1.00	1.00	1.00	108
ك	0.98	0.98	0.98	110
ل	0.97	0.97	0.97	114
م	0.98	0.96	0.97	111
ن	0.99	0.96	0.97	120
ه	0.91	0.97	0.94	88
و	0.99	0.94	0.97	118
ي	0.99	0.96	0.94	112
لأ	0.99	0.97	0.98	113
ة	0.96	0.96	0.96	104

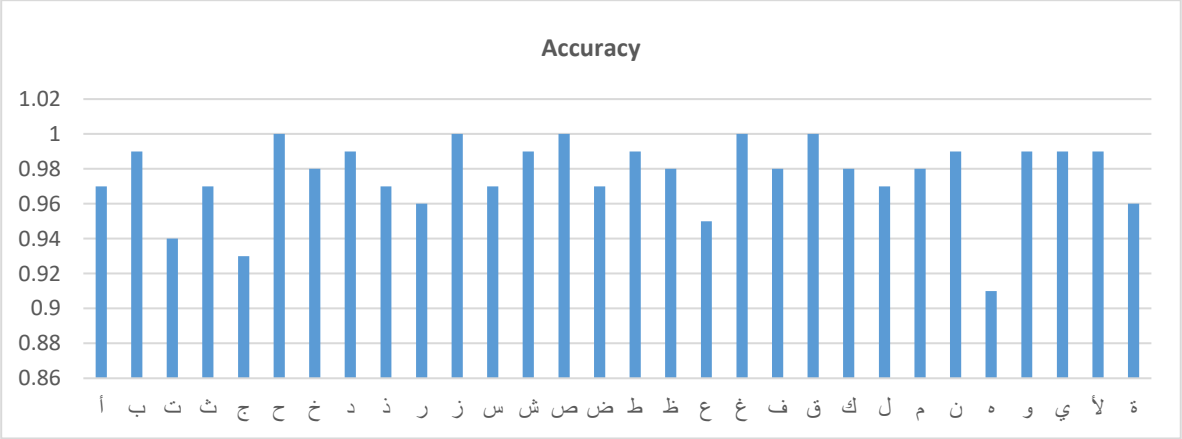


Figure 8: accuracy for each label of 32 classes

For gesture sign language recognition using the media pipe with a traditional machine learning model, we also calculated the accuracy, precision, recall, and F1 score. Additionally, we used the confusion matrix to visualize the model's performance, as shown in Figure .8.

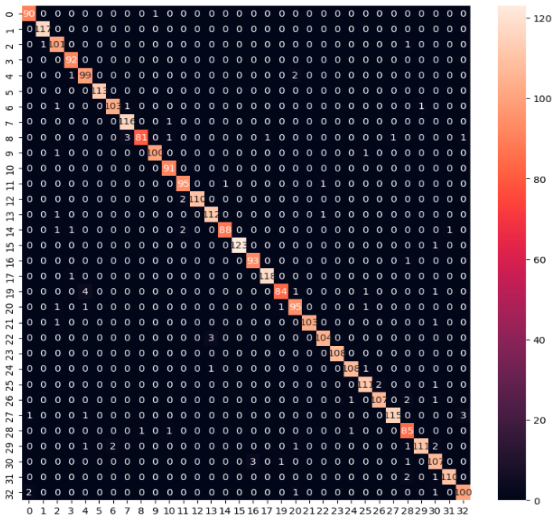


Figure 8. Confusion matrix for sign of Arabic letter recognition

In addition, although the letters د and ذ have the same view and can only be distinguished by the position of the thumb, the system could still recognize them successfully. K-Fold validation was performed on the dataset by taking tenfold. The average accuracy of over ten iterations for different algorithms is shown in Table III. The average can be seen from the accuracy provided that ANN outperforms other machine learning algorithms, such as SVM, and achieves higher accuracy than other algorithms, such as KNN, DT, and RF. For exhaustive testing, each sign language image dataset is preprocessed to extract features using the Media pipe framework and trained in different ML algorithms to classify gestures correctly. An accuracy of 98% is achieved for most of the datasets, which outperform present state-of-art and classify fingerspelling of Sign Languages precisely. A maximum accuracy of 99.50% is gained for Number Sign Language. The testing performance for each dataset is summarized in Table 3.

Table 2: Average accuracy obtained using machine learning algorithms

Dataset	ANN	SVM	KNN	RF	DT
ASL	98%	95%	93%	92%	83%
ASN	100%	99.50%	99.15%	99.15%	99.15%

A. Results of word Recognition

We will train our model to classify 10 words. Each word has 50 videos, and each video has 40 frames. The Sequence Bi_ BiLSTM (bidirectional LSTM) Model consists of two Bidirectional LSTM layers and a Dense layer to the Sequential model.

- The words are عندك كم سنة, محتاج مساعدة, انا تعبان, عامل ايه, تمام الحمدلله, السلام عليكم, اين منزلك, ما اسمك

In our research on Arabic Sign Language recognition, we classified static sign language gestures and move gestures using our dataset. Individual alphabet signs were classified using the media pipe and traditional machine learning models as ANN for fixed sign recognition. At the same time, whole sentences were identified using the Bidirectional LSTM model for move gesture sign recognition. According to our research, using sequence-based models for marker recognition based on time-series prediction significantly differs from using ANN models to process key points in 3D space. In particular, our findings showed that the Bidirectional LSTM models were more effective in recognizing static sign language gestures. In contrast, the media pipe with ANN models was more effective in recognizing static sign language gestures. In the context of Arabic Sign Language recognition, achieving a test data accuracy of 96% is a noteworthy feat. This level of accuracy has been attributed to the efficient use of Bidirectional LSTM and ANN with media pipe models for the recognition of static and dynamic sign language gestures, respectively. An important indicator of the model's ability to learn and generalize patterns from the data is loss minimization and attrition during training. It indicates that the model no longer learns from the training data and can predict the test data accurately.

We used the confusion matrix to visualize the performance of gesture sign words recognition model. The confusion matrix allows for a deeper analysis of its performance.

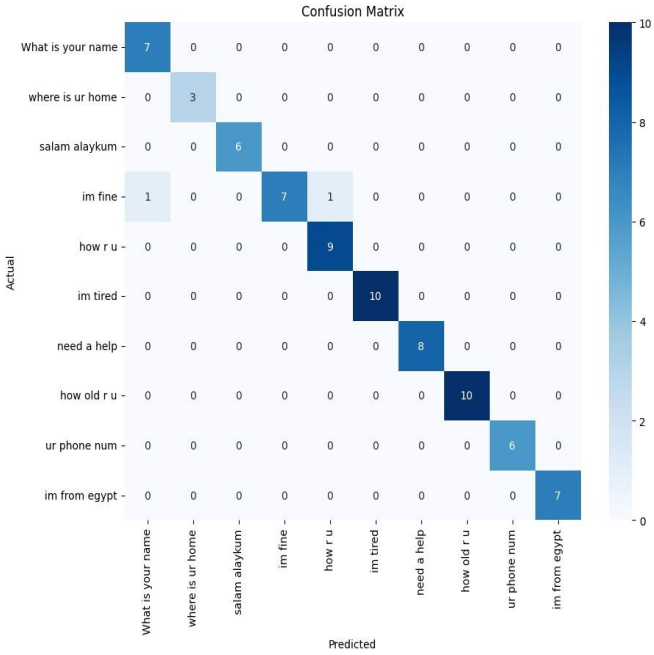
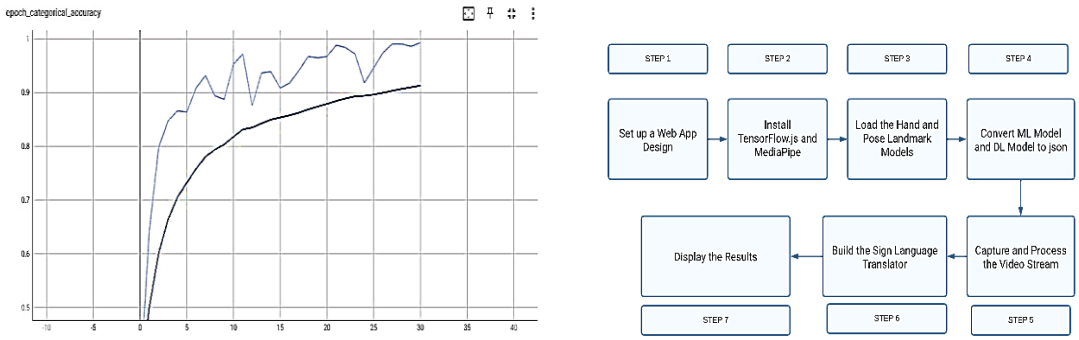


Figure 9. Confusion matrix for sign word recognition

This points us in a clear direction on how separating signals based on their motion can be useful when choosing the correct model for identification. Our experiments achieved an accuracy of approximately 96% for the train data, with low loss and stability after sixty to seventy iterations of training. This indicates that training for sufficient repetitions is critical to achieving optimal results. Our findings will contribute to further technological advances in sign language recognition, ultimately improving communication and outreach to the deaf and hard of hearing. Fig 10 shows an accuracy of 96%.



(a) (b)

Figure 10. (a) accuracy curve of sign words model, (b) . Implementation of ASLT on web page.

First, create a web app using HTML, CSS, and JavaScript. TensorFlow.js is a library for running machine learning models in the browser. Media pipe is a framework for building perceptual computing pipelines. Load the pre-trained hand and pose landmark models provided by Media Pipe using TensorFlow.js. These models are trained to detect the key points or landmarks on the hands and human body. Capture the Video Stream: we use the get User Media API to capture the video stream from the user's webcam. Process the Video Stream: we use TensorFlow.js and Media Pipe to process the video stream and detect the hand and pose landmarks in real time. To build Sequence Once you have the hand and pose landmarks, you can use them to build a sign language detector. You can use a machine learning model to classify the detected landmarks into different sign language gestures. Then, display the detected sign language gesture on the web app as shown in Fig 11.

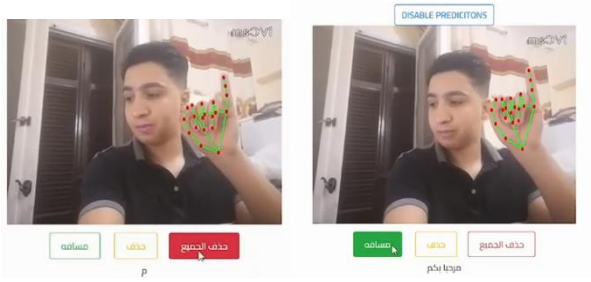


Figure 11 video stream from the user's webcam

In this paper, we developed an efficient and powerful real-time Arabic Sign Language letter, Arabic Sign Language Word recognition model, and number model using media pipe and BiLSTM. The proposed model was developed with the complete Arabic alphabet, which allows users to write full articles using Arabic Sign Language letters in real time. This system gives an accuracy of 98%. It allows everyone to use it through online group video calls. The proposed system will automatically detect hand sign letters and pronounce the result in Arabic using media pipe and BiLSTM. a model. This system gives an accuracy of up to 96%. While the numbers model reaches 100%.

Table 3: Evaluation of the proposed model in relation to existing models

Reference	**Data set type**	**Classifier**	**Accuracy**
[23] (2021)	There are a total of 54,049 images in ArSL2018, representing the 32 ArSL alphabets and signs contributed by 40 signers	CNN	97.29%
[24] (2019)	28 Arabic letters and digits from 0 to 10 are represented in all of the 7,869 images	CNN	90.02%
[25] in (2023)	7,057 images for recognizing 28 Arabic letters	CNN with Media pipe	97.1%

Our model	There are a total of 54,049 images in ArSL2018, representing the 32 ArSL alphabets and signs contributed by 40 signers. Also The dataset has 1000 different permutations of a single alphabet. This makes a total of 32,000 numpy array, and digits from 0 to 10 and 10 common words .	BiLSTM	Arabic letters 98% Digis from 1 to 9 100% Words model 96%

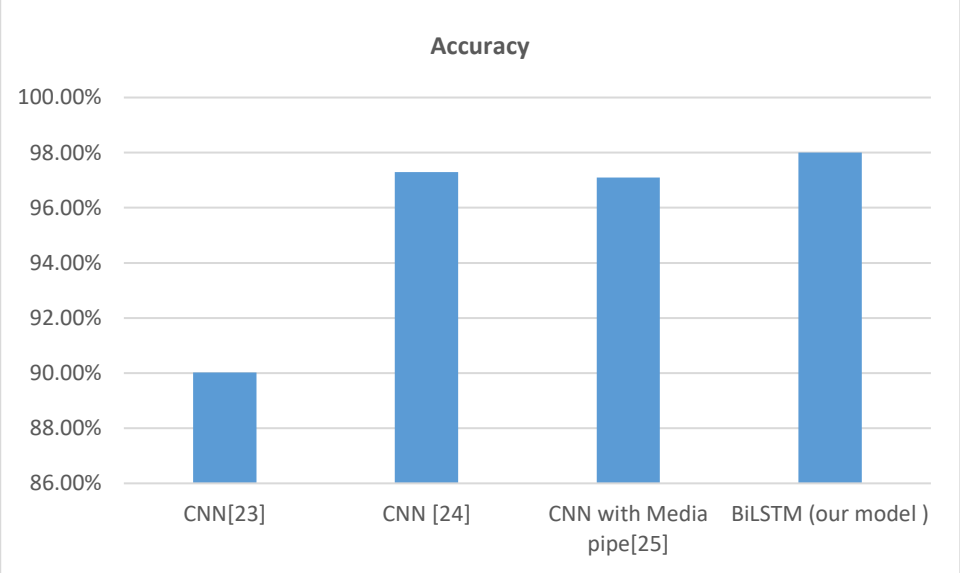


Figure 12: Evaluation of the proposed model in relation to existing models

VII. Conclusion

This study uses media pipe to demonstrate a real-time recognition model for the Arabic Sign Language Alphabet (ASLT)). The major goal was to create a solution that could be used by everyone. Three models were proposed. This work uses ASLT for letters, words, and numbers. It allows everyone to use it through online group video calls. The proposed model was developed with the complete Arabic alphabet, which allows users to write full articles using Arabic Sign Language letters in real time. This study uses six machine learning algorithms: ANN, DT, SVM, RF, KNN, and BiLSTM . ANN and BILSTM achieved the best result, 98%.

References

[1]. M.A. Jalal, R.Chen, R.Moore, et al., "American sign language posture understanding with deep neural networks," in *Proc. of 21st International Conference on Information Fusion (FUSION)*, pp. 573-579, 10-13 Jul 2018. Article (CrossRef Link)
[2]. S.P. Becky, "Sign Language Recognition and Translation: A Multidisciplined Approach From the Field of Artificial Intelligence," *Journal of Deaf Studies and Deaf Education Advance Access*, Vol. 11, no.1, pp.94-101, 2006. Article (CrossRef Link) P. R. Graves and T. A. J.
[3]. M. Mustafa, "A study on Arabic sign language recognition for differently abled using advanced machine learning classifiers", J. of Ambient Intelligence and Humanized Computing 12, 4101–4115, Mar. 2020.
[4]. American Sign Language, National Institute on Deafness and Other Communication Disorders. http://www.nidcd.nih.gov/health/hearing/asl.asp
[5]. Rahib H. Abiyev, "Facial Feature Extraction Techniques for Face Recognition," Journal of Computer Science, Vol.10, no.12, pp.2360-2365, 2014. Article (CrossRef Link)
[6]. C. Jennings, "Robust finger tracking with multiple cameras," in Proc. of Int. Workshop on Recognition, Analysis, and tracking of Faces and Gestures in Real-Time Systems, 1999.Article (CrossRef Link)
[7]. S. Malassiotis, N. Aifanti, and M. G. Strintzis, "A Gesture Recognition System Using 3D Data," in Proc. of IEEE 1st International Symposium on 3D Data Processing Visualization and Transmission, June 2002. Article (CrossRef Link)
[8]. B. S.Parton, "Sign Language Recognition and Translation: A Multidisciplined Approach From the Field of Artificial Intelligence," Journal of Deaf Studies and Deaf Education, Vol.11, no.1, pp.94-101, 2006. Article (CrossRef Link)

[9]. G. Fang, W. Gao, X. Chen, C. Wang, and J. Ma, “Signer independent continuous sign language recognition based on SRN/HMM,” in Proc. of International Gesture Workshop, Gesture and Sign Language in Human-Computer Interaction, pp. 76-85, 2001. Article (CrossRef Link)
[10]. Tharwat, A.; Gaber, T.; Hassanien, A.E.; Shahin, M.K.; Refaat, B. Sift-based arabic sign language recognition system. In Afro-
[11]. European Conference for Industrial Advancement, Proceedings of the First International Afro-European Conference for Industrial Advancement AECIA 2014, Addis Ababa, Ethiopia, 17–19 November 2015; Springer International Publishing: Cham, Switzerland, 2015; pp. 359–370.
[12]. F.N.H. Al-Nuaimy, Proc. 2017 Int. Conf. Eng. Technol. ICET 2017 2018-Janua, 1 (2018).
[13]. B.G. Lee and S.M. Lee, IEEE Sens. J. 18, 1224 (2018).
[14]. H.S. Kala, S. Sushith Rai, S. Pal, K. Uzma Sulthana, and S. Chakma, Proc. - 2018 Int. Conf. Des. Innov. 3Cs Comput. Commun. Control. ICDI3C 2018 97 (2018).
[15]. [Kamruzzaman, M.M. Arabic Sign Language Recognition and Generating Arabic Speech Using Convolutional Neural Network. Wirel. Commun. Mob. Comput. 2020, 2020, 3685614. [CrossRef]
[16]. [Mustafa, M. A study on Arabic sign language recognition for differently abled using advanced machinelearning classifiers. J. Ambient. Intell. Human Comput. 2020, 12, 4101–4115. [CrossRef]
[17]. G. Latif, N. Mohammad, R. AlKhalaf, R. AlKhalaf, J. Alghazo and M. Khan, "An Automatic Arabic Sig Language Recognition System basedon Deep CNN: An Assistive System for the Deaf and Hard of Hearing", International Journal of Computing and Digital Systems, Vol.9, No.4, pages 715-724, Jul. 2020.
[18]. Batool Yahya AlKhuraym, Mohamed Maher Ben Ismail and Ouiem Bchir, “Arabic Sign Language Recognition using Lightweight CNN-based Architecture” International Journal of Advanced Computer Science and Applications(IJACSA), 13(4), 2022. http://dx.doi.org/10.14569/IJACSA.2022.0130438
[19]. C. Wang, H. Mark Liao, Y. Wu, P. Chen, J. Hsieh and I. Yeh, ” CSPNet: A New Backbone that can Enhance Learning Capability of CNN” , 2020 IE E E /C V F Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2020. Available: 10.1109/cvprw50498.2020.00203 .
[20]. [19] K. Simonyan and A. Zisserman, ”Very Deep Convolutional Networks for Large-Scale Image Recognition” , 2014
[21]. A. Bochkovskiy, C. Wang and H. Mark Liao, ”YOLOv4: Opti mal Speed and Accuracy of Object Detection” , 2020. Available: https://arxiv.org/abs/2004.10934
[22]. Zhang, F., Bazarevsky, V., Vakunov, A., Tkachenka, A., Sung, G., Chang, C. L., & Grundmann, M. 2020. MediaPipe Hands: On-device Real-time Hand Tracking. arXiv preprint arXiv:2006.10214
[23]. M. M. Kamruzzaman, "Arabic Sign Language Recognition and Generating Arabic Speech Using Convolutional Neural Network", Wireless Communications and Mobile Computing, vol. 2020, Article ID 3685614, 9 pages, 2020. https://doi.org/10.1155/2020/368561
[24]. A. A. Alani and G. Cosma, “ArSL-CNN: a convolutional neural network for Arabic sign language gesture recognition,” Indonesian journal of electrical engineering and computer science, vol. 22, 2021.
[25]. Ahmad M. J. AL Moustafa, Mohd Shafry Mohd Rahim, Belgacem Bouallegue, Mahmoud M. Khattab, Amr Mohmed Soliman, Gamal Tharwat, Abdelmoty M. Ahmed, "Integrated Mediapipe with a CNN Model for Arabic Sign Language Recognition", Journal of Electrical and Computer Engineering, vol. 2023, Article ID 8870750, 15 pages, 2023. https://doi.org/10.1155/2023/8870750
[26]. https://developers.google.com/mediapipe/solutions/vision/hand_landmarker