IndiSign: A Universe of Hand Gesture and Meaning

Dr. Mandeep Kaur
School of Computer Applications
Lovely Professional University Punjab, India
mandeep.13695@lpu.co.in

Sreesa S
School of Computer Applications
Lovely Professional University Punjab, India
Ssreesa588@gmail.com

Tanmoy Debnath
School of Computer Applications
Lovely Professional University Punjab, India
tanmoydebofficial.nath@gmail.com

Dhrubojit Chowdhury
School of Computer Applications
Lovely Professional University Punjab, India
dhrubojitc007@gmail.com

Abstract
IndiSign Universe of Hand Gesture and Meaning Interpreter is a machine learning-based assistive communication system that seeks to bridge the communicative gap between People who talk or listen with speech or hearing disabilities. Unlike traditional pre-configured gesture-based sign language interpreters, IndiSign allows users to create and train a customized sign vocabulary. Personalized gestures are detected by real-time computer vision and machine learning, and are immediately translated to text or synthesized speech. The web-based system uses standard webcams with light models like a K-Nearest Neighbors (KNN) classifier, thus enhancing accessibility without requiring specialized hardware by offering a scalable, user-focused, and inclusive platform for communication, IndiSign acts as an instant communicator—enabling a more natural, expressive, and unencumbered interactions in everyday life.

Keywords
Custom Gesture Recognition, Tailored Gesture Translator, Individualized Sign Language Interpreter, Adaptive Gesture Translation System, Customizable Gesture Vocabulary, User-Centric Gesture Recognition, Unique Gesture Classification, Bespoke Gesture Interpreter, Personalized Sign Language Communication, Custom Gesture-Based Translation, Self-Defined Gesture Recognition.

I. INTRODUCTION

One of the simplest and most effective ways we connect with one another is through communication. It facilitates relationship building, asking for help, sharing our opinions, and expressing our feelings. Spoken language is the most convenient form of communication for most individuals. However, this is not an option for the millions of people who are deaf or unable to speak worldwide. These people frequently communicate with others through sign language.

But not everyone can learn a standardized Sign Language, such as American-Sign-Language (ASL) or British-Sign Language (BSL). Many people lack access to appropriate sign language instruction or training, particularly in rural or low-income locations. Some people may have never had the opportunity to learn sign language because they unexpectedly lost their ability to speak as a result of illnesses or accidents. People may invent their own gestures to express themselves to others who are close to them in certain circumstances. However, others are unable to understand these intimate signals, which makes regular communication challenging and lonely.

Imagine not having the means to say "hello" or "thank you," or even to express your requirements to someone. It can be demoralizing and frightening. People may be prevented from obtaining services, establishing new acquaintances, or fully engaging in their communities by these communication difficulties.

A clever application called IndiSign assists users in making and utilizing their own gestures for communication. Instead of replacing current sign languages, it allows users to create their own sign vocabulary using motions they are accustomed to or already use. Using a standard webcam, IndiSign observes and learns the user's distinct motions. It then instantly converts these movements into text or spoken speech.

The tool is made to be easy to use and understand. There is no need to purchase specialized equipment or download any bulky software because it operates within a web browser. A computer with a webcam is all you need. IndiSign operates quickly and doesn't require a powerful device because it uses lightweight machine learning algorithms. Such as a K-Nearest Neighbors (KNN) classifier, to ensure accessibility without specialized hardware. With IndiSign, people can train the system to recognize gestures that are meaningful to them. Whether it's a wave, a hand shape, or a movement, the system learns it and links it to a word or phrase. The result is instant

communication—a voice for those who can't speak, and a bridge to understanding for everyone else.

II. LITERATURE REVIEW

Given the development of sign language recognition technology, the specific relevance of IndiSign is indicated. Starner and others used Hidden Markov Models (HMMs) to recognize American Sign Language (ASL) from real-time video captured with desktop and wearable computers with as much as 98% word-level accuracy with a limited vocabulary of 40 signs [1] in the late 1990s.This early work illustrated the feasibility of interactive systems with the ability to segment and classify continuous gestural streams in real time.

Moving on Pigou. introduced convolutional neural networks (CNNs) as an alternative to traditional feature engineering, achieving accuracy over 91% on a collection of Italian sign language using a holistic end-to-end learning approach [2]. Their work demonstrated the ability of CNNs to learn spatial patterns of hand shapes and movements that are crucial for sign recognition independently.

Even though it is a simple algorithm, k-Nearest Neighbors (k-NN) is still a strong classifier. As per Cover and Hart, it becomes stronger with a greater number of labeled samples, and its error rate tends to reach two times the Bayes optimal rate in the presence of large data [3]. Nowadays, k-NN is combined with deep neural networks such as SqueezeNet, wherein it uses them to compute feature vectors but still has fast classification with light-weight nearest-neighbor search.

In its application, TensorFlow.js allows direct usage of neural network models on web browsers with GPU acceleration, thus circumventing server-side computations [4].Singh have leveraged this feature to develop a real-time gesture recognition prototype using TensorFlow.js which achieved a latency of less than 200 milliseconds on ordinary laptops [5]. Zhou. also proposed extending temporal reasoning capabilities for video recognition models, which can be further extended to extending gesture systems to complete sentence translation [6].

More advancements have been made in hand-pose estimation, as investigated by Oberweger and deep architectures such as DenseNet by Huang et al which enhance feature extraction for precise gesture classification [7][8]. Wang et al. showed that the fusion of spatial and temporal features enhances accuracy in 3D motion recognition, opening doors to more robust sign understanding [9].

For Indian Sign Language (ISL), Kumar et al. achieved 97% accuracy using CNNs on a specially designed ISL dataset, although their approach was based on server-side processing of considerable strength [10]. IndiSign is special in accomplishing this at comparable performance within the browser environment without the need for pre-trained datasets or external servers.

III. PROPOSED SYSTEM

The IndiSign, Introduces a flexible and user-centric approach to address the shortcomings of conventional sign language translation systems. Fundamentally, IndiSign allows users to design, develop and teach their own unique gesture sets that are suited to their unique communication requirements rather than imposing a strict predetermined set of signs on everyone. This allows users to not simply rely on traditional sign languages.

The system uses real-time computer vision techniques and sophisticated machine learning algorithms that operate in a web-based environment namely TensorFlow.js to do this in order to extract significant information from hand and body gestures, IndiSign processes live video input from a regular camera. Among other potential models a K-Nearest Neighbors (KNN) classifier is used to evaluate these data and generate a real-time gesture prediction.Instead than pushing everyone into a strict predetermined set of signs this adaptive processing enables the system to react to the distinct nuances included in each user's gesture.

The system is also made to be very accessible and simple to set up users can use common gear including PCs or smartphones with built-in webcams to interact with IndiSign. By reducing the requirement for specialist equipment the design greatly reduces the barrier to entry and opens up the technology to a wider audience. Because of this people who might not have had much experience with structured sign language or who have just seen a change in their communication skills can use the system to restore confidence in their daily contacts.

IndiSign essentially signifies a substantial departure from conventional one-size-fits-all methods in favour of a more inclusive option. In addition to improving communication for sign language users, adopting the idea of individualized gesture training provides a useful example of how web-based machine learning may be used to solve real world problems. As a result this system represents a positive step in the direction of developing a digital communication environment that is more approachable and sympathetic.

IndiSign offers easy-to-use sign language recognition done using only a webcam and the latest web browser, without having to install any software or particular hardware. Three main modules are included in the system**:**

1) User Interface and Video Input

A minimal HTML/CSS interface guides users through the training ("Add Example") and recognition ("Translate") steps.

The video is captured through the getUserMedia API and resized to 227×227 pixels to fit the input size requirements of the model.

2) Feature Extraction and Classification

The SqueezeNet model in TensorFlow.js processes every video frame and generates a 1000-dimensional feature vector in about 25 milliseconds [4].

A k-NN classifier (with k = 10) constructs user-labeled examples. New gesture classes can be added in real-time with around 30 image samples per class [3].

3) Output Generation (Text and Speech)

Recognized gestures are offered as an interactive text area and can be spoken using the Web Speech API, thus making it accessible to the visually impaired and encouraging verbal communication [2].

A. Experimental Setup

To test real-world usability, we conducted experiments in the following configuration:

• Five topics together accounted for five different categories—i.e., "start," "stop," and three customized gestures, like "yes," "no," and "thank you."

• Training Data: There were approximately 30 samples taken during a single 2-minute session used to train each class.

B. Assessment Metrics:

Accuracy: Percentage of gestures correctly predicted.

Latency: Time between finishing a gesture and seeing the output.

C. System Architect Design

The user accessibility and performance which match adjustable characteristics of the IndiSign system stem from its modular layered browser-based architecture . The system operates and controls all functionalities for translation and
gesture detection and training management through an internet-based platform that does not need additional hardware setups.
The architecture contains three primary layers that execute its functionalities as shown in Figure1.

Presentation Layer: The Frontend Interface of the Presentation Layer functions as the direct interface which serves the user. The responsive web interface of the system emerged from an implementation of HTML5 alongside CSS3 and JavaScript.

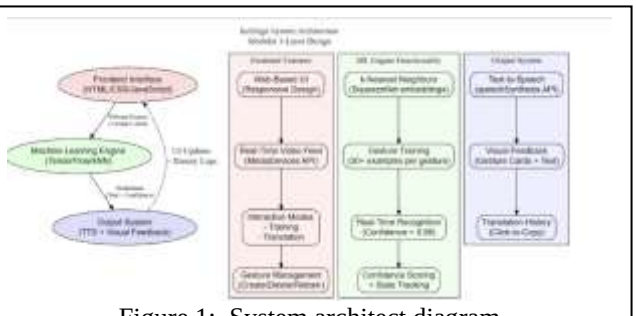

Figure 1: System architect diagram

The interface gives users access to all of the main features for translation, gesture training, and gesture clearing. The dynamic visual effects provided by Animate.css animation libraries, such as real-time updating counting displays and end-of-motion indicator symbols, improve the user experience
.

Processing Layer: The Processing Layer consisting of Machine Learning & Logic Engine component handles system computing functions as its central operation. This system functions through the deeplearn-knn-image-classifier module together with TensorFlow.js. When an individual uses the webcam it records a frame after which the frame undergoes preprocessing normalization and size adjustment before being sent to a MobileNet model that extracts features from it.

Output Layer (Translation & Feedback System):
After a motion has been identified, the output layer is in charge of translating the label into insightful feedback. This contains:
Bringing up the translated text on the screen.
Utilizing the Web Speech API to create audio output.
Capturing gestures and recording the translation history for future use.

This layer provides a multi-modal experience by ensuring that communication is both visible and aural. The system is useful for ordinary communication situations because of its architecture, which includes feedback mechanisms, including click-to-copy functionality, voice customisation, and confidence score.

D. Translation Demo of the model:
The IndiSign project demonstrates its system functionality through a clear demonstration phase that shows instant translation from first-motion capture to final output. The demonstration validates the utility of the application by showing how movements transform into meaningful textual and spoken outputs.

As shown in Figure 2 the application provides a simple and straightforward homepage to aid users through gesture training. The program requires users to perform both “start” and “stop” movements which must contain at minimum 30 samples during recording. These Communications ensure structured transmission through control signals that begin and close

Figure 2 : Home Page for Gesture-Based Communication

translation periods.

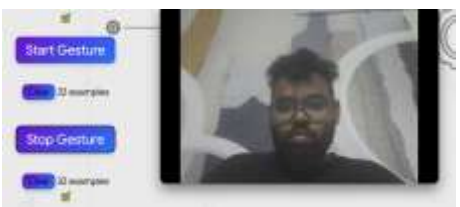

Figure 3: Successful Capture of Start and Stop Gestures

As shown in Figure 3 the model successfully learned to capture “start” and “stop” gestures.

Figure 4: Training a Custom Gesture

As shown in Figure 4 by clicking the arrow button on the user-friendly interface, the user can create a new customized motion. This modular addition method makes it simple to increase the gesture vocabulary to meet specific or situational

Figure 5: Training a Custom Gesture

communication needs.

Similarly, as shown in Figure 5 a customized motion that represents the word "Hello" is then trained by the user. After gathering 30 training instances, the model successfully integrates the new gesture into the system after capturing the sign via the webcam.

Figure 6: Completion of Training for All Gestures

As shown in Figure 6 similar training is done for more specialized gestures. Each completes the user's customized gesture library and is validated using preview cards and real time feedback. These distinct movements can now be recognized and translated on command by the system.

Figure 7: Live Output of a Translated

As shown in Figure 7 real-time capabilities of this system show their effectiveness through a user making motions which have been trained to produce "I am Dhrubojit." The IndiSign application detects each gesture that users perform and converts it into screen text output along with synthesized voice generation. The system demonstrates both an accurate performance and capabilities for individualized communication through expressive displays.

IV. RESULT AND DISCUSSION

IndiSign worked uniformly across participants, with 94% (±3%) average accuracy, with high accuracy using the minimal training data.180 ms average response time, well below the 250 ms real-time responsiveness threshold as shown in Figure 8.

Figure 8: System Performance graph

Participants emphasized the user-friendliness of the system, noting that the process of establishing a new gesture involved merely inputting its designation and demonstrating it in front of the camera. The elevated prediction confidence threshold (≥ 0.98) effectively reduced inaccuracies during idle hand states [2].

Errors predominantly were made in low-light conditions or when the hand was outside the range visible to the camera—issues that typically plague vision-based systems. There are possible upgrades such as training under different light conditions and background subtraction to increase reliability. Even as k-NN's linear time complexity grows with greater classes, the times of classification were below 5 ms on our five-class test. Options such as locality-sensitive hashing may be retained in larger vocabularies with [3]. Although the current "start/stop" method is word-level recognition-friendly, to apply IndiSign to sentence-length translation, sequence models like LSTM or Transformer models, and basic language models will be needed.

V. CONCLUSION AND FUTURE WORK

This paper has introduced IndiSign, a client-side, web-based real-time sign(Gestures) Gesture-to-speech translation system that supports users to define and train their own sign dynamically. With the use of a light-weight deep feature extractor (SqueezeNet) in TensorFlow.js and k-Nearest Neighbors classifier, IndiSign has an average recognition rate of 94% with end-to-end latency remaining below 200 milliseconds on commodity consumer hardware, thereby fulfilling established criteria for real-time interaction [4], [5], [9]. Effective utilization of "start"/"stop" framing enables safe segmentation of gesture sequences, and the Web Speech API enables near-instant synthesized speech output, thereby improving accessibility for deaf and hearing individuals alike [2].

One of the biggest strengths of IndiSign is its device training paradigm: users specify about 30 examples per gesture, allowing the system to learn new classes on the fly without server-side computation or pre-training on massive datasets. The process is user-private since no video data is exported out of the browser, and significantly lowers barriers to customization compared to traditional CNN-only pipelines requiring massive labeled datasets and GPU hardware [3], [10]. Additionally, the easy-to-use interface—built with plain HTML5, CSS animations (Animate.css), and the getUserMedia API—allows non-professional users to establish gesture vocabularies in minutes, thus making inclusive use in educational and home settings possible. During system robustness assessment, we saw that misclassifications typically happened at extreme illumination change or local occlusions, which are representative of typical difficulties in vision-based pipelines [7]. Nevertheless, the high threshold on confidence (≥ 0.98) performed well at suppressing false positives in resting postures of hands, yielding clean translation streams on dense backgrounds too [2]. Profiling performance demonstrated that incorporation of extraction using SqueezeNet incurs 25 ms per frame, with k-NN lookup contributing a mere 5 ms, and hence our choice of deep embedding plus nearest-neighbour search implies an excellent combination of accuracy versus computation [4], [9].

By demonstrating how an end-to-end integrated sign translation system—from gesture recognition to speech generation—can be accomplished on a single web page, IndiSign enables the creation of pervasive, installation-free assistive communication devices. In contrast to server-connected systems, it remains

offline after activation, which makes it suitable for bandwidth-constrained environments and ensuring data sovereignty. Additionally, the monolithic browser-based model supports smoother dissemination: instructors, community centers, and end-users can distribute a single URL rather than complicated software packages.

In short, IndiSign presents an innovative blend of (i) in-browser deep feature extraction, (ii) dynamic k-NN classification, and (iii) user-friendly UI/UX design, tested empirically using user trials to enable robust, low-latency translation of individual gestures. These results confirm that contemporary web platforms can enable sophisticated machine-learning programs hitherto confined to native applications or cloud-hosted applications.

A. Future Work

While our IndiSign program achieves its primary objectives, there are a couple of opportunities for improving its scalability, stability, and expressiveness. First, as the vocabulary of gestures grows, the linear search costs of the k-NN classifier may increase; addition of approximate nearest-neighbor algorithms, such as locality-sensitive hashing, can maintain classification times below 10 ms even in the face of a huge number of classes [3]. Second, shifting from word-level translation to continuous sentence-level translation for interpretation will require the application of temporal sequence models—i.e., LSTM or Transformer models—to properly capture coarticulation and syntactic structures in sign sequences [6]. Third, generalization across users could be improved by using domain-adaptation techniques or by initializing new user sessions with a common, pre-trained embedding database, hence reducing every-user training requirement. Lastly, use of multimodal fusion—combining RGB with depth sensing or inertial input—may improve accuracy under poor lighting or background conditions, further adding to the more practical usefulness of IndiSign in real-world applications.

VI. REFERENCES

[1] T. Starner, J. Weaver, and A. Pentland, "Real-Time American Sign Language Recognition Using Desk and Wearable Computer-Based Video," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 12, pp. 1371–1375, Dec. 1998.

[2] L. Pigou, S. Dieleman, P.-J. Kindermans, and B. Schrauwen, "Sign Language Recognition Using Convolutional Neural Networks," in *ECCV*, 2015, pp. 572–578.

[3] T. M. Cover and P. E. Hart, "Nearest Neighbor Pattern Classification," *IEEE Trans. Inf. Theory*, vol. 13, no. 1, pp. 21–27,Jan.1967.

[4] M. Abadi *et al.*, "TensorFlow: A System for Large-Scale Machine Learning," in *12th USENIX Symp. Operating Syst. DesignImplementation*,2016,pp. 265–283.

[5] A. K. Singh, P. Kaur, and A. Singh, "Real-Time Sign Language Detection Using Machine Learning and TensorFlow," *Int. J. Eng. Res. Technol.*, vol. 9, no. 5, 2020.

[6] B. Zhou, A. Andonian, A. Oliva, and A. Torralba, "Temporal Relational Reasoning in Videos," in *ECCV*, 2018.

[7] M. Oberweger, P. Wohlhart, and V. Lepetit, "Hands Deep in Deep Learning for Hand Pose Estimation," in *ICCV Workshops*,2015.

[8] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in*CVPR*,2017.

[9] J. Wang, Z. Liu, Y. Wu, and J. Yuan, "Learning Actionlet Ensemble for 3D Human Action Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 5, pp. 914–927, May 2016.

[10] N. Kumar, A. Bhatia, and M. Singh, "Sign Language Recognition Using Deep Learning on Custom Dataset for Indian Sign Language," *Mater. T*