SIGNIFY: A Pre-Primary Learning and Practice Hub for Indian Sign Language

Suraj Dora
Member
Pillai College of Engineering
New Panvel, Maharashtra – 410206, India

Aditya Gupta
Member
Pillai College of Engineering
New Panvel, Maharashtra – 410206, India

Jyotirmay Prasad
Member
Pillai College of Engineering,
New Panvel, Maharashtra – 410206, India

Vasudha Kadam
Member
Pillai College of Engineering
New Panvel, Maharashtra – 410206, India

K.S. Suresh Babu
Guide
Pillai College of Engineering,
New Panvel, Maharashtra – 410206, India

ABSTRACT
Signify is an interactive e-learning platform that has an influence on Machine Learning (ML) and Deep Learning (DL) to make Indian Sign Language (ISL) learning easier for pre-primary students. The system combines advanced hand detection models like MediaPipe and OpenCV to isolate hand gestures from the camera feed with accuracy, which ensures precise recognition.

A convolutional neural network (CNN) handles these gestures in real time, allowing sign-to-text and text-to-sign conversions while giving quick feedback through interactive quizzes. Also, an analytical dashboard keeps track of student progress offering insights into learning patterns and unlocking extra resources as students move forward.

By joining deep learning with real-time hand tracking, Signify promotes inclusivity, boosts communication skills, and backs a structured engaging approach to sign language education.

General Terms
Sign Language, Deep Learning, Hand Gesture Recognition, Computer Vision, Interactive Learning.

Keywords
Sign Language, Deep Learning, Hand Gesture Recognition, Computer Vision, Interactive Learning.

1. INTRODUCTION
Sign language plays a vital role as a bridge for smooth communication within the Deaf and Hard of Hearing (DHH) Community, and Accessibility to Appropriate Language Education is one of the persisting challenges faced by individuals with hearing impairments. Sign language plays a pivotal role in bridging the gap between Deaf and Hard of Hearing People and the rest of the community to bridge that communication gap. But due to the dependence on traditional ways of teaching, staleness in resources, and lack of trained teachers, structured tools for learning Indian Sign Language (ISL), especially for young learners, the availability is negligible.

The recent rise in Artificial Intelligence and Deep Learning has promised significant improvements in the field of Sign Language Education providing a real-time interactive learning solution. The study describes Signify, an AI-based e-learning system that fills this gap by hooking up computer vision-based recognition of hand gestures through an atmosphere appropriate for education and engagement. It uses MediaPipe and OpenCV to track hands in real-time and a Convolutional Neural Network to classify gestures in an interactive learning environment for the child to practice signs of numbers and letters from A to Z. Besides recognition, it gamifies learning experience, applies adaptive learning, and has an analytical dashboard to track student progress. This appropriate approach provides data-driven insights for the teacher and the parent to facilitate learning interventions that are very specific. The exposure of Signify to ISL at an early age promotes linguistic inclusivity, accessibility, and engagement. Therefore, to make it more inclusive, it enhances the learning of sign language and makes it more engaging and scalable.

It shall evaluate its technical architecture and implementation and shall discuss its pedagogical impacts. Describing the role of Signify as a platform that triggers AI-driven policies to bridge communication gaps and empower young learners.

2. LITERATURE SURVEY
2.1 Literature Review
In the recent past, developments in AI and deep learning have been changing learning technology with more efficiency, making it possible to learn in a personalized and interactive environment. One such innovation is AI-driven sign language learning platform that offers personalized, interactive learning mechanisms for promoting linguistic accessibility and early exposure to sign languages. With the significance of Indian Sign Language (ISL) for building bridges among the Deaf and Hard of Hearing (DHH) community, it is becoming apparent that there is a more urgent need for intelligent systems in enhancing the education of ISL by recognizing hand gestures in real-time.

In this survey of literature, an attempt has been made to articulate the existing AI-driven systems of learning sign languages, covering the methodology, technical frameworks, learning models, and user engagement strategies. Thus, the future work will try to identify the potential challenges and limitations of existing systems in the development of the ISL learning process by the analysis of the technology used for gesture recognition and the interaction of students with the system. The present study ultimately has a goal which includes participation in the production of e-learning platforms that are more precise, scalable,, and engaging, like Signify, bridging accessibility and inclusivity gaps in sign language education.

Liang et al. [1] focuses on a comprehensive survey of techniques for sign language recognition by the rule- based method of traditionalists, statisticians, and current deep learners. The merit of their work comes from the fact that it is broad and critically informs the reader with a deep insight into some techniques, like sequence modeling and encoder- decoder frameworks. Such benchmarking is disadvantaged due to the problem of non-availability of standard databases and evaluation protocols, as well as a problem with diversity in the data to make such findings generalizable. It does not, however, state how accurately it performs with numbers, but says it performs well in terms of qualitative aspects of how robust and applicable the model is.

Jiang et al. [2] presents a novel machine translation framework that tries to bridge spoken and signed languages in a machine translation system based on SignWriting as an intermediary form of representation. The advantage of this paper is using SignWriting in a new way to provide a structured representation that merges textual and sign language modality within the framework of sequence-to-sequence modeling with the attention mechanism. Such an implementation has the disadvantage in respect of its work primarily due to the lack of SignWriting datasets of high quality and the difficulty in capturing very nuanced features in sign language. Explicit accuracy numbers are not reported, the paper reports results in terms of translation fidelity and effectiveness in bridging the language modalities that the framework achieves.

Zhou et al. [3] proposed a direct sign language translation model without referential gloss annotation via large-scale vision-and-language pretraining. The one possible upside of the system is very easy end-to-end translation with which much better generalization over diverse sign expressions can be achieved. Drawbacks are such that it may be too hard to grasp some fine points of sign language without gloss annotations and much depends on both the quality and quantity of the dataset used for pretraining. This work does not present exact numbers of accuracy but gives a more qualitative description of improved generalization and reduced error propagation as evidence of success.

Chen et al. [4] delves into a very simple baseline of multi-modal transfer learning for sign language translation. They use transfer learning to combine linguistic features with visual information, it greatly simplifies the training process, plus the tasks can be used for multiple translation tasks based on a very strong foundation. Only verified models are reused, enabling it to significantly improve translation performance. The primary drawback of this methodology though is the availability and domain applicability of these pre-trained models, respectively. There is hardly any numeric value in its evaluation, which is also highly qualitative. But these limitations may be regarded as stepping stones for just further work into multi-modal fusions in the translation of sign languages.

Yin et al. [5] describes the development of MLSLT, a general framework for developing multilingual sign language translations. This system involves using deep learning techniques in the form of sequence-to-sequence models with attention in extracting language-agnostic features to enable translation among various sign languages. Although not without flaws, the key benefit of the approach is its ability to merge representations from very distinct source languages as a precursor to extended multilingual application. It faces the primary limitation of the low availability of high quality, multilingual data and the difficulty in capturing language-specific details. Furthermore, the evaluation depends mostly on qualitative comparison and includes limited quantitative benchmarks. This represents an important step in the direction of a universal sign language translation system; therefore, we take it into account.

Sharma et al. [6] has implemented a full translation pipeline for converting spoken language into Indian Sign Language via natural language processing. The system provides an integrative environment for speech recognition, text translation, and sign language synthesis while positioning it to support improved communication for the Deaf and Hard of Hearing community. Its main benefit lies in the errorless bridging from speech to sign, thus greatly enhancing real-world communication potential. On the downside, this system is highly error-prone because errors, particularly from the part of the speech recognizer, are propagated as there are not enough numerical measures to evaluate its performance accurately. This shortcoming notwithstanding, the approach may be considered an optimistic way of bridging the gulf between spoken and sign languages.

Duarte et al. [7] explores How2Sign, a large-scale multimodal dataset for continuous American Sign Language that compiles video and audio data synchronized at the textual level. Whereas its advantage is both on an extensive scale and multimodal diversity that give a robust foundation for more accurate and standardized models, complexity in annotating continuous sign language may lead to inconsistency hence this paper does not report direct numerical accuracy for translation tasks but only provides a critical resource for future evaluation. The dataset will, nevertheless, become one main pioneer toward improvements in research on continuous sign languages.

Roelofsen et al. [8] looks into extending sign language translation techniques to the domain of healthcare. The work focuses on retraining deep learning models for more domain-specific vocabulary and contextual requirements in medical communication, thereby increasing translation accuracy within healthcare settings. The paper has to offer the development of a context-aware system for improved clarity in medical communication. However, it could be considered more domain-specific in approach and, therefore, find restrictions regarding generalization; besides, the empirical data is more qualitative and thinly backed with numerical data. In this respect, however, the paper puts forward an encouraging domain-specific solution- communication in healthcare to and from the Deaf involving any gestures.

Kahlon and Singh [9] conducted a systematic review of machine translation techniques from text to sign language. They very methodically take a look into a broad spectrum of approaches— right from rule-based to statistical and deep learning methods— which offers an overall comprehensive view of the field. Their review gains strength from structured synthesis calling out prevailing trends and gaps (such as standardized datasets and consistent evaluation metrics) but then again, this is a review and does not have new experimental results to contribute and its assessments are qualitative more than anything else. Still, it provides a useful guidepost to researchers who want to make improvements in text-to-sign language translation systems.

Zheng et al. [10] presents the upgrade of a sign language translation model, explicitly explainable for coping with the translation of lengthy sign sentences. The model adopts state-of-the-art deep learning architectures in converting directly from extended sign sequences into text, yet within supplemented mechanisms for explaining what it "thinks." The primary benefit is the control of longer sequences with transparency for the translations which altogether build trust from the end-users. The increased model complexity, however, calls for relatively higher computational resources and longer training periods, besides the evaluation being mostly qualitative with only a small portion of numerical accuracy data. It is with such challenges that the work marks a significant step closer to building translation models that are interpretable yet effective.

He [11] delves into new neural architectures and focused training methodologies that cater to the special requirements under sign language data conditions. Thus, its merit lies in the development of model designs that provide very good baselines for translation systems. Only small data sizes and simple evaluation strategies have restricted scalability and general accuracy. It is in fact this very paper that can be considered as an encouraging base, to be later fine-tuned with more strong datasets and advanced evaluation methods.

2.2 Summary of Literature Survey

A literature review is an objective, critical summary of published research literature relevant to a topic under consideration for research. The summary is presented here.

Table 1. Summary of Literature Survey

Paper Name	Observation
Sign Language Translation: A Survey of Approaches and Techniques [1]	Advantage: Provides an extensive and critical survey of various sign language translation techniques—including rule- based methods, statistical models, and modern deep learning approaches (such as sequence modeling and encoder–decoder frameworks). This comprehensive overview is beneficial for researchers entering the field and sets a solid foundation for future work by identifying trends and promising techniques. Disadvantage: The survey is limited by the absence of standardized datasets and evaluation protocols, making direct quantitative comparisons challenging. Additionally, issues related to data diversity and representation are not fully resolved, which may impact the generalizability of the surveyed approaches.
Machine Translation between Spoken Languages and Signed Languages Represented in SignWriting [2]	Advantage: Introduces an innovative framework that leverages SignWriting as a structured intermediary to bridge spoken and signed languages. The use of sequence-to-sequence modeling with attention mechanisms allows for effective integration of textual and visual modalities, paving the way for unified translation across language modalities. Disadvantage: The framework is hindered by the limited availability of high-quality SignWriting datasets, which constrains its scalability and generalizability. Furthermore, the method struggles to capture fine-grained linguistic nuances and cultural variations inherent in sign language, and the evaluation is primarily qualitative with few quantitative accuracy metrics provided.
Gloss-free Sign Language Translation: Improving from Visual-Language Pretraining [3]	Advantage: Smooth end-to-end translation that jumps over the intermediate gloss annotation. Generalization and efficiency in translating visual sign inputs into textual output directly, on the cover, as the model is able to achieve this just come out much better when going all-in with large-scale visual-language pretraining. This approach also takes away some potential error propagation of gloss generation. Disadvantages: It might fail to capture some details of sign language that are quite subtle and context-specific without explicit gloss annotations. Another issue can arise from the fact that it heavily depends on the quality and diversity of the datasets in the pre-training phase. Finally, the authors have only provided qualitative measures of performance without citing a single numerical accuracy metric.
A Simple Multi-Modality Transfer Learning Baseline for Sign Language Translation [4]	Advantage: It adopts a simple and effective baseline that combines visual and linguistic modalities very well through transfer learning. In that it uses the pre-trained model, it hurts less in terms of how hard it is to train, but it makes a very strong baseline that is much better for sign language translation. Such integration of multimodality proves the possibility of reusing already existing models that are going to capture visual gestures along with text representations. Disadvantage: The proposed approach completely favors the availability and compatibility of pre-trained models in the certain characteristics of data for sign languages. The evaluation, however, majorly speaks qualitatively with a few limited numbers of accuracy metrics with respect to numbers, not relating to temporal dynamics and subtle nuances of gestures.
MLSLT: Towards Multilingual Sign Language Translation [5]	Advantage: Introduces a multilingual translation framework that leverages advanced sequence-to-sequence models with attention mechanisms. The method is designed to extract language-agnostic features, enabling translation across different sign languages. This approach holds promise for bridging language modalities by unifying representations from diverse sign language inputs. Disadvantage: The effectiveness of the

	model is constrained by the scarcity of high-quality, multilingual sign language datasets, which limits its ability to generalize across various languages. Moreover, the evaluation is predominantly qualitative, lacking extensive quantitative benchmarks, and it struggles to address the fine-grained linguistic nuances and cultural variations inherent to different sign languages.
Translating Speech to Indian Sign Language Using Natural Language Processing [6]	Advantage: We introduce a multilingual translation framework, which is actually composed of advanced sequence-to-sequence models with attention mechanisms. The method is meant to extract language-free features that would enable translation for different sign languages. It has the potential to unify representations across diverse sign language inputs for bridging language modalities. Disadvantage: More limitations come to place due to the scarcity of high-quality multilingual datasets of sign languages. It does not cope with much generalization but otherwise is described by the author as being positive. Besides quality weighing, this model still has a lot of problems that relate to different types of linguistic features and sub-cultural elements of a sign language.
How2Sign: A Large-scale Multimodal Dataset for Continuous American Sign Language [7]	Advantage: Offers a large-scale, comprehensive, multimodal dataset of continuous American Sign Language. The resource provides synchronized video, audio, and textual annotations, making it very valuable for training and benchmarking systems that work on sign language recognition and translation. The diversity and scale of this dataset will indeed provide a really strong basis for researchers to be able to develop models more accurately and in a standardized manner. Disadvantage: Inconsistency in the annotation is another factor that can indicatively lead to a loosely modeled annotation effect on the learning model. Since this is a study based on datasets and does not directly come to show any accuracy metric for translation or recognition tasks, thus additional preprocessing might be needed to fully captivate the temporal dynamics and subtle details of the gestures inherent in continuous sign language.
Sign Language Translation in a Healthcare Setting [8]	Advantage: Utilizes translation techniques within a medical environment, building upon the preprocessing of domain data along with the adaptations of models to realize quality translations in context, quite ideal for medical communication. Disadvantage: The specific nature of the application leaves little room for general use in more general signed language applications. Data scarcity and the absence of standardized metrics for evaluation in a healthcare setting presents even more difficulty since the efficiency is largely discussed qualitatively without an in-depth numerical analysis.
Machine Translation from Text to Sign Language: A Systematic Review [9]	Advantage: This paper introduces a comprehensive compilation of all the methodologies that exist for text-to-sign language translation, thus also providing a bird's eye view of the various approaches (rule-based, statistical, and deep learning). It brings out all trends and best practices in this area. Disadvantage: It is not a novel system, not new experimental data. Its evaluation is primarily qualitative, pinpointing the persistent challenges and what is to be fully addressed in future research; among the general persistent challenges discussed is the need for standardized datasets and consistent quantitative metrics.
An Improved Sign Language Translation Model with Explainable Adaptations for Processing Long Sign Sentences [10]	Advantage: It puts forward an innovative translation model meant for the long sentences of sign language, incorporating explainable adaptations that increase transparency and trust to users as well. It is this approach that successfully deals with some of the challenges in processing extended sequences concerning sign languages. Disadvantage: High computational demands and long training times, with obvious qualitative improvement but low detailed numerical accuracy metrics sometimes might remain challenges in capturing context-dependent fine nuances inherent in continuous sign language.
Research of a Sign Language Translation System Based on Deep Learning [11]	Advantage: Shows that applying deep learning architectures to sign language translation is possible, with some new model designs and training strategies to get a good performance baseline. This paper contributes some precious insight into the early-stage system development that can be useful for sign language translation. Disadvantage: The system will work just okay on small datasets and that too is mostly okay under preliminary evaluation protocols. So this could well imply that it might be far from scalable and generalizable. The work shall not have been brought up to fully leverage the state-of-the-art in deep learning and the quantitative performance metrics are sparse which means they need further refinement and validation.

2.3 Inference

The breakthroughs in e-learning platforms which are integrated with machine learning (ML) and deep learning (DL) for education in sign languages have placed notable improvements on real-time hand gesture recognition and interactive learning experiences. Literature reviews demonstrate well the performance that models like MediaPipe and OpenCV bring relative to accurate tracking of hand movements. Convolutional neural networks further make possible the translation accuracy of sign-to-text and text-to-sign. This greets an innovation to offer more exciting and adaptive learning environments, especially with young learners.

Yet, a lot of conspicuous gaps continue to exist regarding the personalization and adaptability of such systems when it comes to addressing the varied learning needs of pre-primary students. Solutions that are available today do not quite meet the criteria of adaptive difficulty scaling, personalized learning pathways, and feedback on the progress made by each student- at least not yet. What is even worse is the fact that very little research has examined how to prod these multimodal cues, for example, facial expressions and situational gestures, aspects fetched from the mainstream sign language community into the system.

These gaps would significantly enhance the effectiveness of ISL learning platforms. Personalized learning trajectories provide dynamic feedback, but additionally, AI-analyzed gesture displays assist both in accessibility and retention for early learners. Additionally, driving AI-based analytics to track patterns of engagement and optimize content for contextual delivery would make for a more structured and inclusive approach to the pedagogy of sign language for the long run to instill language within participants and create their communicative competence.

3. PROPOSED SYSTEM

Signify is an interactive e-learning system that will educate pre-primary students through teaching Indian Sign Language. This will use the hand detection model, i.e., MediaPipe, and OpenCV with CNN to transform hand gestures accurately into corresponding text and vice versa. In the implementation, the solution will interact with young learners through interactive quizzes accompanied by prompt feedback while an analytical dashboard keeps track of their progress by adapting content for improved learning experiences.

3.1 Existing System Architecture

The existing system [3] utilizes an end-to-end encoder–decoder architecture that is optimized for direct gloss-free translation of sign languages. The architecture maps sign language video sequence inputs directly to outputs in text, bypassing the need for any intermediate gloss annotations and hence making the translation process more efficient. Major components collaborate with each other within the framework to perform translation at a high level of accuracy.

Figure 3(a) shows the architecture of the integrated GFSLT model, emphasizing the encoder-decoder structure on real-time continuous sign language video processing. The encoder is responsible for extracting deep visual features from the input video stream using a series of 2D CNN layers. On the other hand, the decoder, which is based on the Transformer model, works on generating the corresponding sequence of text. What is indeed remarkable about this design is that it skips the conventional gloss-based intermediary step; hence, holding back the before-informed information channel bottleneck habitually attached to gloss annotations.

The second detailed view is for the Vision Embedding module. In this module, each frame of the sign language video goes through multiple 2D CNN layers over which robust spatial features are extracted. Next, these features are aggregated and passed on to the subsequent layers of the network to ensure that the visual representation to be created is rich in detail as well as being highly discriminative.

Figure 2 now provides an illustration of the Visual-Language Pretraining (VLP) approach in the system. In this approach, the masked self-supervised learning pre-training phase combines with CLIP-based contrastive pretraining to align visual and textual representations in a shared semantic space. Due to this reason, the model initializes well and translates much better overall than any method that uses just glosses or just text—for the first time ever, achieving stunning BLEU-4 scores (≥+5 on the PHOENIX 14T dataset and ≥+3 on the CSL-Daily dataset).

Taken together, the diagrams and performance metrics prove the robustness of the GFSLT architecture. This is an end-to-end workflow that mirrors the process by which Signify processes hand gestures into meaningful textual output while delivering a superior translation performance lagging only baseline always-on low-latency performance by a meager 1.33x.

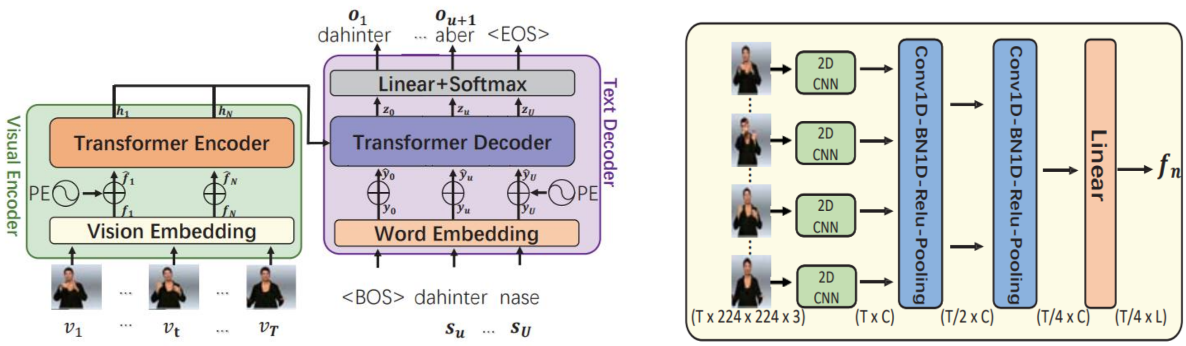

Figure 1: Existing system architecture [3]

3.2 Input and Output Specifications of Existing System

3.2.1 Input Specification
The input accepts continuous sign language video sequences, standardized to a uniform resolution and frame rate for optimal feature extraction. [3]

3.2.2 Output Specification
An output produces real-time, fluent textual translations of the sign language input via its encoder–decoder framework.[3]

3.3 Proposed System Architecture
The proposed system consists of the following elements:

3.3.1 Dataset
Between 400 and 500 images per ISL gesture making all 26 letters (A-Z) and 10 numerical digits (0-9) were taken to ensure wide coverage. Images were taken at varying light conditions, angles, and backgrounds to guarantee all-roundness. Every image is annotated down to its corresponding gesture label. Annotations will help specifically define the training process and enhance the accuracy of recognition.

3.3.2 Feature Extraction
MediaPipe and OpenCV detect landmarks, and from images and video frames, spatial and temporal features can be extracted from hands. To ensure the model is robust, there should be techniques applied such as normalization, resizing, and data augmentation (rotation, flipping, addition of noise).

3.3.3 Classification
This CNN is meant for extracting detailed spatial features from each frame. Sequence modeling will be applied to ensure the successful achievement of continuous sign recognition by modeling temporal dependencies

3.3.4 Feature Database
A dedicated database stores the pre-extracted and annotated features from the training dataset. This repository facilitates rapid matching during live gesture recognition, ensuring real-time responsiveness.

3.3.5 Similarity Measurement
Compare live query features against the repository. Recognition results are assigned confidence scores to ensure high accuracy even in challenging conditions. Recognized gestures and their corresponding meanings are instantly displayed on the user interface, with visual and auditory feedback to enhance learning.

3.3.6 Virtual Interaction
A web-based interface and a Progressive Web App provide intuitive, accessible access for pre-primary students.

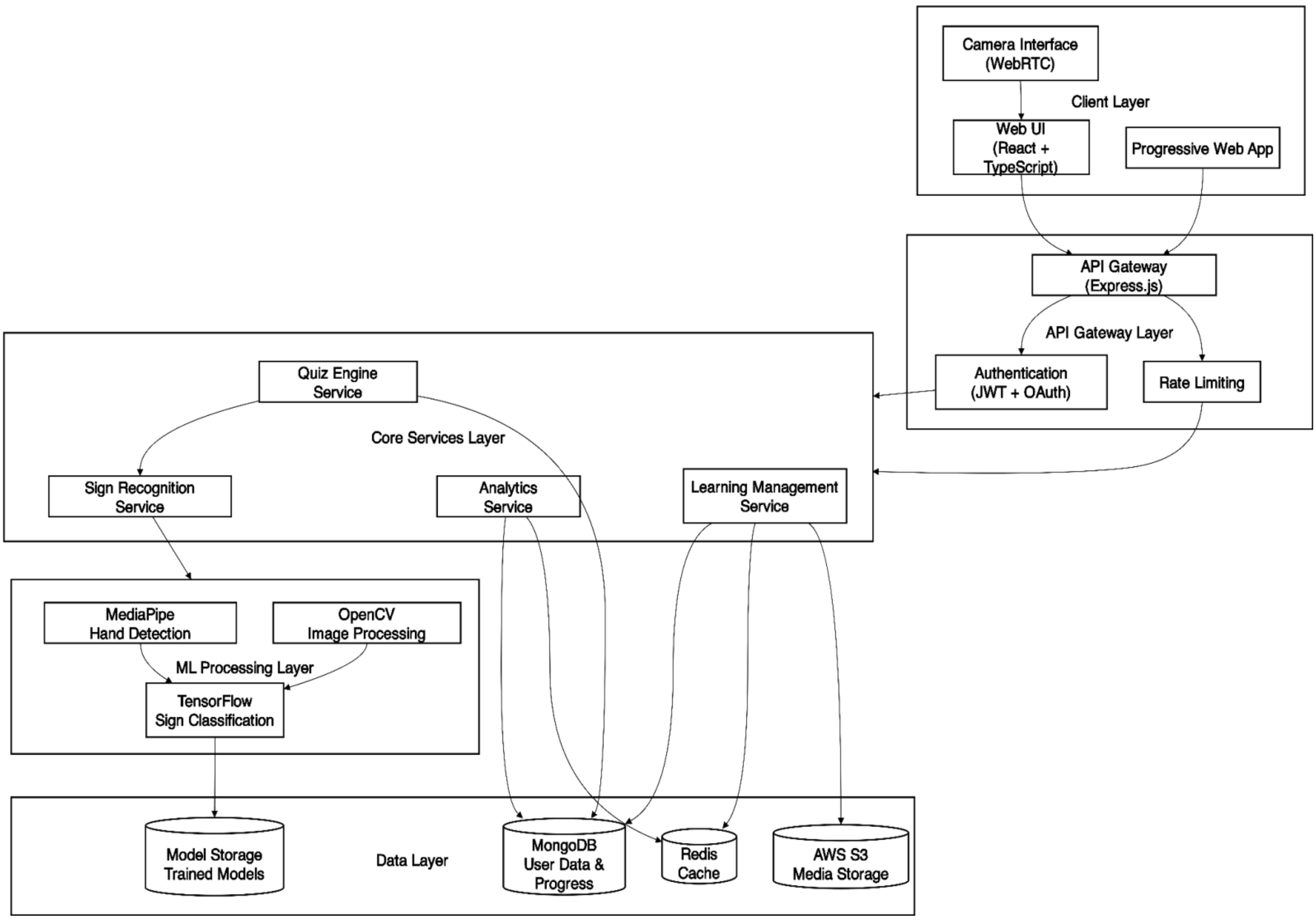

Figure 2: Proposed system architecture

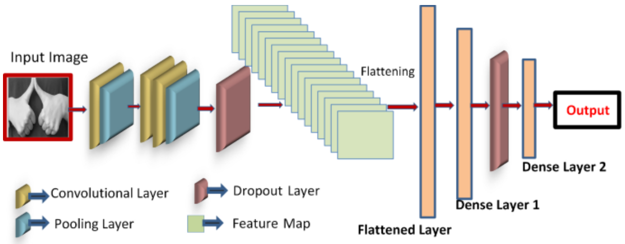

Figure 3: Block diagram for Feature Extraction[15]

3.4 Datasets Used

Table 4. Dataset Details

Dataset Used	No. of Images	Types of Images	No. of Categories
Signify ISL Dataset	15,000–20,000	Hand gestures for A–Z and 0–9 in Indian Sign Language	36

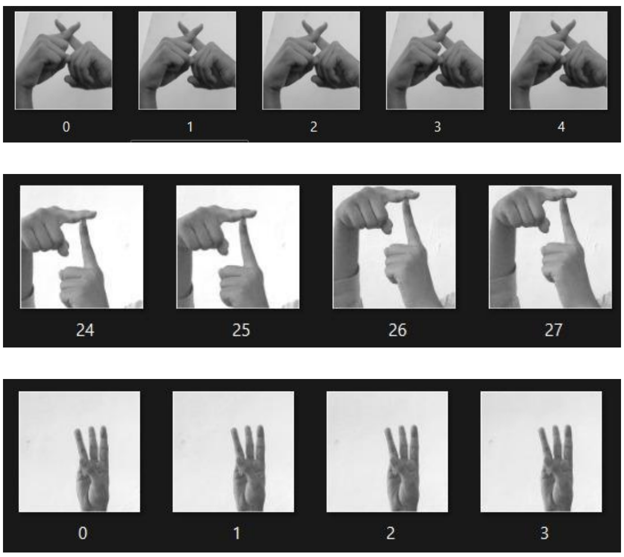

Figure 4: Sample of Dataset

3.5 Input and Output Specifications of Proposed System

3.5.1 Input Specification

During quiz-based interactions, a user can use their camera to provide real-time images or videos for the system to recognize hand gestures involved with ISL letters and numbers. An alternative for the user is to display an image and then identify the corresponding alphabet or number.

3.5.2 Output Specification

The system shall determine whether the user has performed the correct gesture or not and provide appropriate feedback in real-time. It shall predict the gesture performed and display the recognized letter or number, thus making the learning process interactive and effective. Alongside, it will also identify whether the alphabet or number matches with the input image displayed.

3.6 Use Case

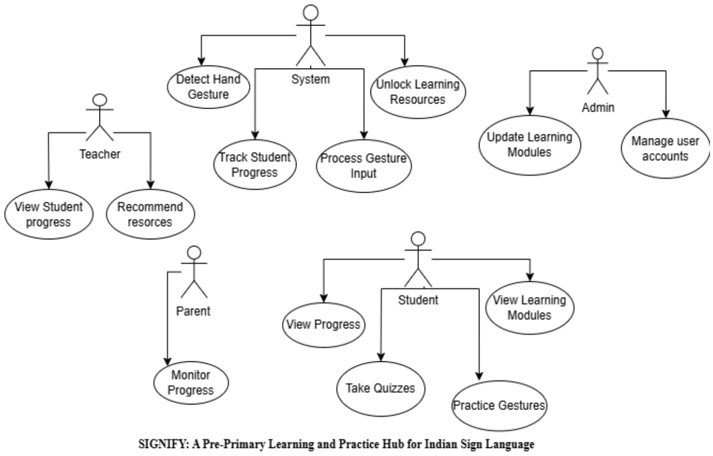

Figure 5: Use Case Diagram

4. METHODOLOGY

In this section, we present the proposed methodology for the SIGNIFY platform. The system is supported by a CNN model for real-time hand gesture recognition, using MediaPipe and OpenCV for feature extraction. The platform will ensure feedback on sign language gestures at the time of gesture-making, so that an interactive learning environment can be realized.

4.1 Data Collection

The collection of the SIGNIFY dataset will begin with collecting ISL gesture data of alphabets (A-Z) and numbers (0-9) from multiple users, taken under different lighting conditions, from multiple angles, and in front of different backgrounds. Each sign will be accompanied by metadata containing the actual sign in text form. This, combined with the structured dataset at such an early stage, will lay down steps for building the recognition model.

4.2 Data Preprocessing

The other thing to do before model training is to standardize the images in the dataset. On resizing images to a fixed dimension of 224x224 pixels, the next step in preprocessing is to normalize pixel values to a range of 0 to 1 grayscale. To further enhance general model performance, grayscale conversion is done for pixel values in a range of 0 to 1. It reduces computational complexity. Finally, Data augmentation techniques, including rotation, flipping, and addition of noise, are applied to improve model generalization. All these steps shall bring uniformity to the input data and make the model more robust.

4.3 Feature Extraction with CNN

For the purpose of feature extraction, this model implements the use of a Convolutional Neural Network due to its high efficiency in the capture of patterns and hierarchies regarding position on an image. This architecture constitutes several layers that work together to process input gestures and extract meaningful features. In more specific terms, convolutional layers "spot" significant gestures, hand contours, and finger positions while pooling layers maintain the essential features and reduce the spatial dimensions. The final mapping of these extracted features to the associated ISL gestures would be done by fully connected layers, and the softmax layer, in return, provides probabilities for each class identifying the actual gesture.

4.4 Cosine Similarity

After the feature vectors are extracted, Signify employs cosine similarity to compare hand gestures and identify patterns among similar gestures. Cosine similarity is calculated as:

$$Cosine \quad Similarity = \frac{A \cdot B}{||A||||B||}$$

where A and B denote the feature vectors of two gestures, and $||A||$ and $||B||$ represent their magnitudes. A higher cosine similarity score indicates that the gestures are visually similar in terms of the extracted features. By computing cosine similarity between the gesture features captured from the user and those stored in the dataset, the system determines the most accurate match.

4.5 Model Training and Fine-Tuning

Fine-tuning the CNN model performance meant for recognition of ISL gestures. The process starts with pre-training the model on a rather large dataset of images, practically ImageNet, to leverage learned representations. The researcher will then proceed to conduct a fine-tuning process on the ISL dataset by implementing transfer learning, thus the model adapts to sign language gestures.

Using the Adam optimizer based on backpropagation and backpropagation optimization to minimize errors in classification will check accuracies further. The performance of the model shall be calculated based on cross entropy loss to classify gestures correctly. The training will be conducted over multiple epochs but shall be applied using the mini-batch gradient descent method in order to come up with a model that generalizes well on new data.

4.6 Evaluation and Performance Metrics

Performance is evaluated using a variety of criteria, including accuracy, precision, and recall. Additionally, user feedback and system interaction are used to fine-tune prediction levels and improve answer relevancy.

4.6.1 Precision

Precision measures the exactness of the gesture recognition system. It is defined as the ratio of correctly recognized hand gestures (true positives) to the total number of gestures the system predicted for a given sign. High precision indicates that when the system predicts a gesture, it is very likely to be correct.

$$Precision = \frac{Number \quad of \quad correctly \quad recognized \quad gestures}{Total \quad number \quad of \quad gestures \quad predicted}$$

4.6.2 Recall

Recall quantifies the system's ability to detect all instances of a particular hand gesture. It is the ratio of correctly recognized gestures to the total number of actual gestures present in the dataset. A higher recall means the system is effective at capturing most of the true gestures without missing many.

$$Recall = \frac{Number \quad of \quad correctly \quad recognized \quad gestures}{Total \quad number \quad of \quad actual \quad gestures}$$

4.6.3 Accuracy

Accuracy reflects the overall performance of the Signify gesture recognition system. It is calculated as the ratio of all correctly recognized gestures (true positives) across all classes to the total number of gestures evaluated. This metric shows how close the system's predictions are to the actual gestures.

$$Accuracy = \frac{Total \; number \; correctly \; recognized \quad gestures}{Total \quad number \quad of \quad gestures \quad evaluated}$$

4.7 Inference and User Interaction

In the process of inference, real hand gestures captured via the webcam of the user are compared against the trained dataset using cosine similarity to detect the gestures. After detecting the gesture, the system classifies the gesture in real-time. The workflow for user interaction should include gesture input where the user carries out a sign in front of the camera; then feature extraction using the CNN model will match it with the trained dataset. Once the system detects the gesture, it provides real-time feedback if the gesture is right or wrong. An analytical dashboard also keeps logs of user progress and recommends exercises based on trends.

Deep learning, real-time gesture tracking, and user analytics will be integrated by the SIGNIFY platform to make sure that better opportunities are given for pre-primary students to master sign languages. This shall make learning a more engaging experience besides maintaining its effectiveness so that young learners can practice and perfect their signing skills with much ease.

5. HARDWARE AND SOFTWARE SPECIFICATIONS

The experiment is carried out using a computer system with the differing hardware and software characteristics listed in Tables 5 and 6, respectively.

Table 5. Hardware Details

Processor	2.40GHz Intel
RAM	16 GB
HDD	512 GB
Microphone	Desktop
Speaker	Desktop
Camera	Webcam/ Desktop

Table 6. Software Details

Operating System	Windows 10
Programming Languages	Python, HTML, CSS, JavaScript
IDE	VS Code
Packages	PIL, NumPy, Keras, TensorFlow, Flask, React
Operating System	Windows 10

6. RESULTS

The Signify system has proven its potential to be implemented to facilitate real-time hand gesture recognition into the Indian Sign Language (ISL) learning system. Users can interact with its system via hand gestures, thereby making the learning process interactive, more informative, and, most importantly, engaging to the learner. Results from the successful implementation reveal the effectiveness of convolutional neural networks in accurately recognizing and conveying meaningful feedback based on gestures.

Figure 6: E-Learning platform SIGNIFY

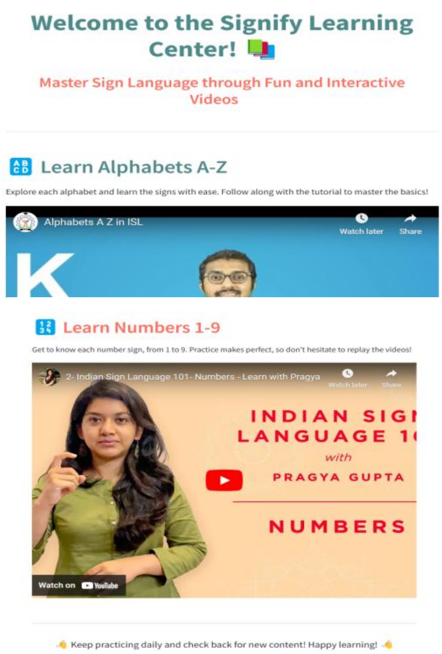

Figure 7: Learn Alphabets and Numbers[13, 14]

Figure 8: Practice quiz

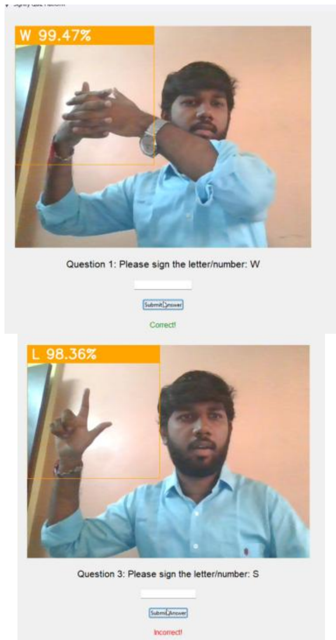

Figure 9: Sign to text quiz

The user will start performing gestures in front of the camera, and the inputs will be preprocessed using MediaPipe and OpenCV for feature extraction. The features extracted will be passed through a pre-trained CNN model; in this model, they will be classified into predefined ISL signs. Then, classified gestures will be compared with a dataset of ISL signs, after real-time feedback can be provided.

Signify is bundled with a method to track progress using logs of student interactions for gesture accuracy and interpretation. Informed by these analytics, the system may recommend personalized ways through which the learner can improve his or her gesture accuracy and comprehension. It integrates real-time recognition and feedback mechanisms to ensure that students are able to practically use ISL.

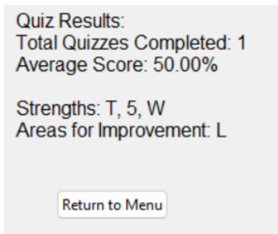

Figure 10: Progress Analytics

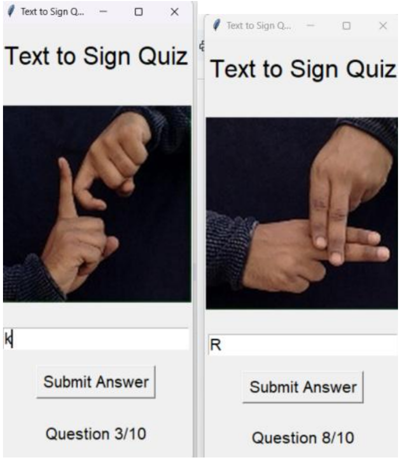

Figure 11: Text to sign quiz

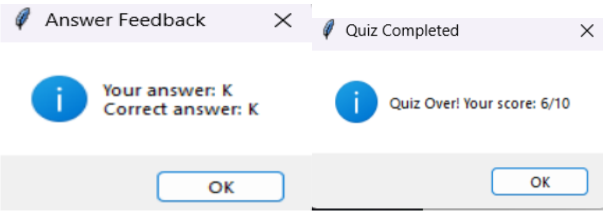

Figure 12: Text to sign practice quiz

Figure 13: Resources Hub for upgrade to Higher Level Learning[12]

7. PERFORMANCE EVALUATION

An assessment metric quantifies the predictive model's performance. This often entails training a model on a dataset, then applying the model to make predictions on a holdout dataset that was not utilized during training and comparing the results to the predicted values in the holdout dataset. Precision and recall are based on a knowledge and assessment of relevance.

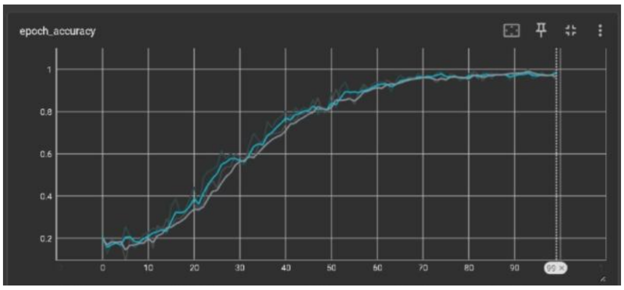

Figure 14: Training Accuracy

Accuracy is the primary metric that states what portion of the gestures is classified correctly. The model is about 95.31% accurate in classifying gestures on the training dataset, while that is 98.35% on the validation dataset (i.e. good for recognition). Another analysis that can be made is loss analysis, to see how much error is there in prediction; lower values in that case mean a more precise model. The model's loss for training data continued to decrease to 0.141 and that for validation data further went down to 0.0198; therefore, the model will be quite good at generalizing unseen data.

Table 7. Summary of Model Performance

Metric	Training Set	Validation Set
Accuracy	95.31%	98.35%
Loss	0.141	0.0198
Precision (avg)	0.97	0.99
Recall (avg)	0.96	0.99
AUC Score	0.98	0.99

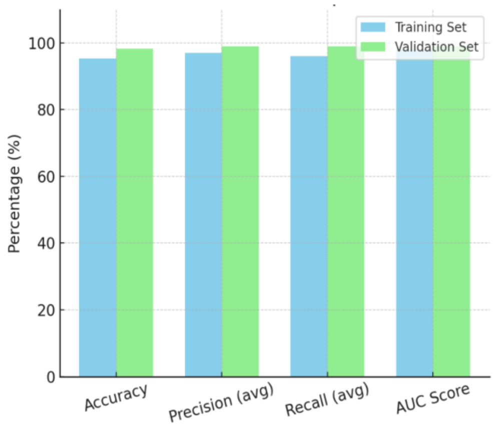

Figure 14: Model Metric Comparison

Figure 15: Loss Comparison

Precision and recall are normally utilized as indicators to decide the success of a classification method, where precision quantifies the actual correctness of the alleged gestures and recall determines how much of the real gestures were identified by the model. The system ensured extremely high values of both precision and recall thereby guaranteeing accuracy in the classification of gestures.

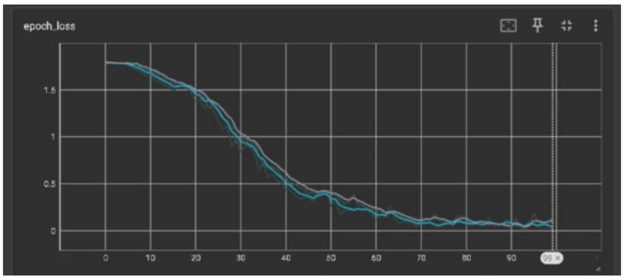

Figure 16: Training Loss

Other details that the confusion matrix provided included visibility into gestures misclassified which eventually lavished refinements on the model training towards greater accuracy of the model. Furthermore, the AUC-ROC curve was used to assess the model's class separability capabilities. The high AUC ratings of 0.98 for the training set and 0.99 for the validation set confirm the model's superior ability to distinguish between different gesture classes.

Beyond accuracy measurements, it is critical to assess performance consistency across multiple evaluation methodologies. The model is stable and has limited volatility between training and validation outcomes, indicating that it is not prone to overfitting or underfitting. These extensive testing confirm the model's excellent efficacy and reliability for gesture recognition tasks.

8. CONCLUSION

Signify fills a considerable gap in the academic resources used by the pre-primary age group for learning ISL. Signify caters to the needs of inclusion and early communication skills by merging real-time hand gesture detection with interactive learning tools. Quizzes; feedback in real-time and on the go; and progress monitoring will make learning fun and successful. Signify is a stride to let sign language education be more approachable and substantial with a spotlight on accessibility and capacity building for young learners.

As we look ahead, Signify's flexible and expandable design sets the stage for big steps forward. Adding datasets in many languages and gesture models for different regions could help it reach people all over the world supporting various language communities and local sign language styles. Also, using 3D hand tracking and depth-sensing tech would help solve problems like hidden gestures and allow the system to recognize more detailed complex signs. These upgrades promise to make the system more reliable and give users a complete, more engaging way to learn.

What's more, Signify doesn't just help Deaf and Hard of Hearing people by making sign language learning easier to access and more organized. It also helps the general public become more aware and better at communicating. By closing communication gaps and encouraging inclusion, Signify has the power to make a big lasting difference in education, social life, and accessibility. This paves the way for a society where people connect and understand each other better.

9. FUTURE SCOPE

9.1Multilingual Support Enhancement

Scaling up Signify to support various sign languages requires the collection of new datasets, development of linguistic variant models, and incorporation of regionally accurate gestures during training to ensure precise translation and interpretation. Expanding multilingual capabilities will not only promote inclusivity but also establish Signify as a globally adaptable learning solution.

9.2 Boosting Gesture Recognition with 3D Models

The combination of 3D hand-tracking helps correct occlusion, depth perception, and complex movements. It will enable identification to improve recognition through properly interpreting complex sign structures and thus enabling a better learning experience.

10. ACKNOWLEDGMENTS

It gives us great pleasure and huge joy to convey our sincere gratitude to Prof. K. S. Suresh Babu, our dissertation project guide, whose unwavering support allowed us to work with enthusiasm. His unwavering drive, tolerance, and outstanding initiative in discussions throughout the dissertation project have helped us to an incalculable degree. We owe him a great deal for his tremendous advice and unwavering support, which enabled us to successfully complete this dissertation on time. The experience of working under his direction has been memorable and productive. He was always available to offer us counsel, assistance, and direction throughout the duration of our project, despite his hectic schedule. The completion of this project would not have been possible without his enthusiasm, patient direction and persistent support.

We would especially like to thank Dr. Sharvari S. Govilkar, the head of the department of computer engineering, for providing us with the chance to work on this significant project, which allowed us to acquire new ideas and apply them.

Lastly, we would like to thank Principal Dr. Sandeep Murlidhar Joshi for providing us with the facilities and opportunity to complete this project.

We owe our friends a huge debt of gratitude for their inspiring, motivating, and encouraging help with our effort.

11. REFERENCES

[1] Z. Liang, H. Li, and J. Chai, "Sign Language Translation: A Survey of Approaches and Techniques," Electronics, vol. 12, no. 12, p. 2678, Jun. 2023.

[2] Z. Jiang, A. Moryossef, M. Müller, and S. Ebling, "Machine Translation between Spoken Languages and Signed Languages Represented in SignWriting," Findings of the Association for Computational Linguistics: EACL 2023, pp. 1232–1243, May 2023.

[3] B. Zhou, Z. Chen, A. Clapes, J. Wan, Y. Liang, S. Escalera, Z. Lei, and D. Zhang, "Gloss-free Sign Language Translation: Improving from Visual-Language Pretraining," in Proc. IEEE/CVF Int. Conf. Comput. Vision (ICCV), 2023, pp. 1234–1243.

[4] Y. Chen, F. Wei, X. Sun, Z. Wu, and S. Lin, "A Simple Multi-Modality Transfer Learning Baseline for Sign Language Translation," arXiv preprint arXiv:2203.04287v2, Mar. 2022.

[5] A. Yin, Z. Zhao, W. Jin, M. Zhang, X. Zeng, and X. He, "MLSLT: Towards Multilingual Sign Language Translation," in Proc. IEEE/CVF Conf. Comput. Vision Pattern Recognit. (CVPR), 2022, pp. 15707–15716.

[6] P. Sharma, D. Tulsian, C. Verma, P. Sharma, and N. Nancy, "Translating Speech to Indian Sign Language Using Natural Language Processing," Future Internet, vol. 14, no. 9, p. 253, Aug. 2022.

[7] A. Duarte, S. Palaskar, L. Ventura, D. Ghadiyaram, K. DeHaan, F. Metze, J. Torres, and X. Giro-i-Nieto, "How2Sign: A Large-scale Multimodal Dataset for Continuous American Sign Language," in Proc. IEEE/CVF Conf. Comput. Vision Pattern Recognit. (CVPR), 2021, pp. 1–10.

[8] F. Roelofsen, L. Esselink, S. Mende-Gillings, and A. Smeijers, "Sign Language Translation in a Healthcare Setting," in Proc. Translation and Interpreting Technology Online Conf. (TRITON), 2021, pp. 110–124.

[9] N. K. Kahlon and W. Singh, "Machine translation from text to sign language: a systematic review," Universal Access in the Information Society, vol. 22, no. 1, pp. 1–35, Jul. 2021.

[10] J. Zheng, Z. Zhao, M. Chen, J. Chen, C. Wu, Y. Chen, X. Shi, and Y. Tong, "An Improved Sign Language Translation Model with Explainable Adaptations for Processing Long Sign Sentences," Computational Intelligence and Neuroscience, vol. 2020, Article ID 8816125, Oct. 2020.

[11] S. He, "Research of a Sign Language Translation System Based on Deep Learning," in Proc. 2019 Int. Conf. Artificial Intelligence and Advanced Manufacturing (AIAM), 2019, pp. 1–5.

[12] Indian Sign Language Research and Training Center (ISLRTC), "ISLRTC Official Website." https://islrtc.nic.in

[13] Pragya Gupta, "Indian Sign Language 101- Numbers - Learn with Pragya," YouTube, [Online Video]. Available:https://www.youtube.com/watch?v=vnH2BmcSRMA

[14] Indian Sign Language Research and Training Centre, "Alphabets A-Z in ISL," YouTube, [Online Video]. Available:https://www.youtube.com/watch?v=VdwKSyza5oI.

[15] Sharma, S., Singh, S. "Recognition of Indian Sign Language (ISL) Using Deep Learning Model," in Wireless Pers Commun 123, 671–692 (2022).