Sign Language Detection Using Computer Vision

Rutuja Moon[1], Vaibhavi Bure[2], Priyanka Telang[3], Prajwal Mohod[4], Rupali Pawade[5], Prof. Apurva Bhuyar[6]
Department of Computer Science, Sipna College of Engineering and Technology, (Sant Gadge Baba Amravati University), Amravati, India.

Abstract—**This project aims to create a system for detecting sign language by utilizing computer vision methods. For those with hearing loss or who are deaf, sign language provides a crucial way to communicate. This approach aims to close the communication gap by translating sign language movements into written text automatically. The suggested solution involves utilizing a camera to capture images of people making sign language gestures, and then processing the image frames to recognize and interpret the hand movements. Algorithms are used to categorize these gestures and associate them with the appropriate words or phrases in sign language.**

In the last ten years, sign language interpretation technology has made remarkable progress, working to resolve communication obstacles experienced by individuals with hearing disabilities. This research proposes a groundbreaking system that converts hand gestures into English text, which is then translated into Hindi. The research employs a hybrid methodology, combining software development, machine learning, and user testing to ensure both technical effectiveness and practical usability of the system. This investigation tackles shortcomings in previous methods by improving the speed and precision of gesture recognition.

Index Terms—**Computer vision, gesture recognition, machine learning, multilingual translation, real-time recognition, sign language detection, speed improvement, technical effectiveness, text conversion.**

I. INTRODUCTION

The paper examines progress in sign language interpretation technology, emphasizing its vital role in overcoming communication barriers for people with hearing impairments or deafness [43]. Sign language serves as a visual method of communication primarily used by individuals with hearing loss or those who are deaf [40]. It relies on a blend of hand gestures, facial expressions, and body movements to convey meaning and facilitate communication. While it is a crucial tool for the deaf and hard-of-hearing community, it poses challenges for those unfamiliar with the language, making communication difficult. Recently, there has been increasing attention toward creating systems for detecting and recognizing sign language, utilizing technologies like computer vision and machine learning [41]. These advanced systems aim to enhance accessibility and interaction by converting sign language gestures into text or speech automatically. Effective communication is a basic human right and essential for all, yet individuals who depend on sign language frequently encounter substantial obstacles when interacting with those unfamiliar with it. This communication gap restricts social integration and access to opportunities, particularly in a linguistically diverse society like India. The research addresses these issues by through an integrated translation component. The system addresses these limitations by incorporating a more extensive vocabulary database, enhanced gesture recognition algorithms, support for Indian regional signs, and an optimized processing framework to reduce delays. Data gathered from usability tests and surveys are examined using both quantitative and qualitative methods, with qualitative feedback being used to improve the user interface and enhance functionality. A comparison with existing systems demonstrates the superior performance and user experience offered by this model, marking a significant advancement in enhancing communication for the hearing-impaired community, this proposal introduces an innovative system capable of interpreting hand gestures into English text and then translating the text into Hindi, using advanced machine learning algorithms and a user-centric design. The problem stems from limitations in existing technologies, including restricted gesture vocabularies, dependency on high-quality video input, and inadequate multilingual support. These factors undermine the effectiveness of communication tools for the deaf community, especially in situations that require dynamic, real-time interactions. This study

seeks to overcome these challenges by enhancing gesture recognition accuracy, expanding vocabulary databases, and incorporating multilingual translation capabilities. This study aims to create a reliable and efficient system for translating sign language that ensures accessibility and inclusivity, improves communication efficiency, and provides practical usability across diverse user groups. The system focuses on realtime gesture recognition, integration of India's national language, and the reduction of environmental dependencies. The importance of this study is rooted in its ability to enhance the lives of the deaf community by fostering inclusivity and increasing access to social participation, education, and employment opportunities. By overcoming the shortcomings of current models and utilizing a holistic mixed-method approach, the study sets the foundation to develop more accessible and advanced tools for communication.

The focus of this study extends to interpreting hand gestures into English text and speech while translating the text into hindi language, with a particular focus on Indian national language. However, the study is limited by factors such as dependency on sensor accuracy, potential challenges in real-time processing for complex gestures, and the system's reliance on predefined gesture vocabularies.

This research offers a thorough solution to the challenges associated with sign language interpretation and translation, with a particular focus on the Indian context. By combining advanced gesture recognition technology with multilingual translation capabilities, the system offers a powerful tool for enhancing communication accessibility.

II. LITERATURE REVIEW

In recent years, the field of sign language translation and interpretation has made notable strides, with researchers delving into various areas to improve communication for individuals with hearing impairments. This thorough literature review examines the ongoing developments in rural sign language translation, progress in live translation technologies, and the role of artificial intelligence and machine learning in facilitating communication for those with hearing and speech challenges [45].

A 2024 study highlighted the unique challenges faced by rural interpreters, including ethical dilemmas and resource limitations that can lead to professional isolation and burnout[1]. To address these issues, studies have employed a multi-faceted approach to defining 'rural' areas, integrating geographic factors, population thresholds, and specific guidance from relevant authorities[1]. This research has identified significant gaps in interpreter education and service delivery in rural settings, which can inform future training programs and support initiatives[1]. However, the absence of historical data on rural interpreting practices limits the ability to develop targeted programming and underscores the underrepresentation of rural interpreters in national assessments of needs.

Recent research has concentrated on developments in real-time sign language translation systems A 2023 survey of progress made in the last five years highlighted the incorporation of Internet of Things (IoT) technology to improve communication for individuals who are deaf or hard of hearing[2]. These advanced systems strive to close the communication gap for individuals who are deaf or hard of hearing by tackling issues like the lack of standardized sign languages and the demand for efficient translators[2]. Researchers have investigated various technologies, such as machine learning, augmented reality, and wearable devices, to enhance sign language recognition and translation[2]. However, several limitations persist, such as the challenges of translating dynamic gestures in real-time, the need for powerful hardware, and the lack of standardization among different sign languages, which continue to complicate the development of universal translation systems[2].

A 2024 study on American Sign Language (ASL) recognition and translation using a wearable inertial motion capture system showed notable advancements in this field[3]. This study gathered a comprehensive dataset of 300 frequently used ASL sentences from three participants, resulting in 7400 data points[3]. The recognition model developed in this study achieved exceptional accuracy at both the word and sentence levels, while the translation model demonstrated promising results with a relatively low word error rate[3]. The recognition model attained remarkable accuracy rates of 99.07% for individual words and

97.34% for full sentences[3]. The conversion model, based on an encoder-decoder structure, had a word error rate of 16.63%[3]. This groundbreaking research showcases the potential of using reliable inertial data for effective sign language recognition and translation, clearing the path for future developments in the field[3].

Research on Indian Sign Language (ISL) has also gained attention, as highlighted in a 2023 multi-level taxonomy review. This study emphasized the lack of adequate research on ISL compared to other sign languages and explored several data collection methods, such as Kinect-assisted, Leap motion, camera-based, and glove-integrated techniques[4]. Researchers have emphasized the challenges in feature extraction and classification accuracy for ISL, noting the limited work on deep learning approaches in this context[4]. This gap in research presents an opportunity for future studies to explore innovative solutions for ISL recognition and translation[4].

In 2024, a thorough review was carried out on technologies such as deep learning (DL), machine learning (ML), and artificial intelligence (AI) aimed at improving communication for individuals who are deaf or mute [5]. Research has investigated several applications, such as speech recognition, sign language interpretation, and text-to-speech synthesis[5]. The potential of these cutting-edge technologies to enhance communication for the deaf or mute community has been emphasized, along with the ethical concerns surrounding the integration of AI into communication tools[5]. This research underscores the importance of balancing technological advancements with ethical considerations to guarantee the creation of communication solutions that are inclusive and accessible[5].

Lastly, a 2023 study on neural text-to-sign language machine translation focused on developing a system to translate text from Brazilian Portuguese into Brazilian Sign Language (Libras)[6]. This ongoing project aims to develop a system to translate text from Brazilian Portuguese into Brazilian Sign Language (Libras), utilizing a realistic three-dimensional avatar for presentation[6]. This innovative research addresses key issues such as establishing a written representation of sign language, constructing parallel corpora, creating NMT models, and visual presentation of translations[6]. The creation of such systems could greatly enhance communication accessibility for individuals who are deaf or hard of hearing in Brazil and could serve as a model for similar projects in other countries and languages[6].

Although these studies show considerable advancements in sign language interpretation and translation, there are still several challenges and limitations to address. These include difficulties in real-time gesture translation, particularly in complex backgrounds and varying illumination conditions, the need for advanced hardware that may hinder portability and accessibility, and the lack of standardization among different sign languages. Additionally, the high costs associated with developing advanced wearable devices and augmented reality systems for sign language translation remain a concern. By addressing the limitations of current technologies, focusing on inclusivity and accessibility, and considering the ethical implications of AI integration, researchers can continue to develop creative solutions that improve communication for individuals who are deaf or hard of hearing. As the field progresses, it is crucial to maintain a balance between technological advancement and the preservation of linguistic and cultural diversity in sign languages, ensuring that these communication tools serve the needs of all users effectively and respectfully.

III. METHODOLOGY

A. Identifying the Problem

Creating a system capable of recognizing and interpreting gestures in American Sign Language (ASL) requires creating a solution designed to detect and translate these movements, it involves several complex challenges, primarily revolving around recognizing the nuanced variations in how individuals use their hands to form gestures and how to map these to text in a meaningful and accurate way. American Sign Language (ASL) is a complete and established language with unique syntax, grammar, and expressions. It is essential for the system to go beyond recognizing individual hand gestures and encompass the complete depth and complexity of ASL, including movements, facial expressions, and body posture. This problem requires not only understanding the gestures themselves but also interpreting them correctly in context, something that is often overlooked in gesture recognition systems. For simplicity and practicality,

initial development could focus on recognizing static ASL alphabet letters.

These letters form the basic building blocks of the language, and if a system can successfully recognize these, it can be extended to more complex signs and eventually complete sentences and phrases. The system must also account for variations in hand shapes, finger positions, and environmental factors that affect the way gestures are performed. Recognizing the correct gesture is just one part of the challenge; the system must also output the corresponding English text or even multiple languages if desired. Additionally, the system should be adaptive to various contexts: different lighting, backgrounds, or even personal idiosyncrasies in how signs are performed. Given these challenges. It becomes apparent that this task involves much more than simple gesture classification—it requires building a robust, context-aware recognition system that is capable of learning from a wide variety of inputs.

B.   Collecting the Dataset

The success of a machine learning model is primarily influenced by the amount and quality of the data used during the training process. For ASL gesture recognition, a comprehensive dataset is paramount. The dataset should contain multiple instances of each of the 26 ASL alphabet gestures performed in various settings, including changes in lighting, hand orientation, and skin tone. A diverse dataset is necessary to train the system to generalize well across different conditions. Fortunately, several public datasets are available, such as the ASL Alphabet Dataset, which provides labeled images of hands forming the ASL alphabet. However, these datasets may not capture the full variability needed for a robust model, and as such, additional data collection may be required. It's essential to include examples of different hand sizes, orientations, and backgrounds, as these factors can significantly affect the recognition performance. For instance, a gesture made in a well-lit room with a neutral background may look vastly different when performed in low light or against a cluttered or patterned backdrop. The dataset should also capture diversity in terms of ethnicity, gender, and age, as hand shapes and skin tones vary widely, and the model must be trained to handle these variations without bias. One effective approach for building a dataset is to capture video footage of people performing the ASL alphabet from different angles,

under various lighting conditions, and with diverse backgrounds. This video data could be segmented into individual frames for each gesture, ensuring that the system can learn both the hand position and the temporal aspect of the gesture (i.e., the movement over time).

Labeling the dataset accurately is another critical step, as mislabeled data can undermine the model's learning process, leading to errors in gesture recognition. To train the model effectively and assess its performance on unseen data, it is essential to partition the dataset into test, validation, and training sets.

C. Preprocessing the Dataset

The data must be preprocessed before being input into a machine learning model to guarantee the data is in an appropriate format for the model's effective use [33]. Data preprocessing is an essential stage, as raw data is often noisy and inconsistent, and without proper handling, it could lead to inaccurate predictions. One of the first steps in preprocessing is resizing the images to a consistent shape. Images in the dataset may vary in size, and feeding them into a neural network with varying dimensions can cause problems. By resizing all images to the same size (for instance, 224x224 pixels), we ensure that the input is standardized. Additionally, images must be normalized. Normalization involves adjusting pixel scaling values to a defined range, like 0 to 1 to minimize large intensity variations that may interfere with the model's ability to learn effectively. An important preprocessing step is data augmentation, which seeks to artificially increase the dataset by applying transformations such as rotation, scaling, translation, and flipping. These methods improve the model's consistency in recognizing gestures, no matter the variations in position or orientation, ultimately boosting its performance under various conditions.

Another preprocessing step involves noise reduction and image filtering. Some images in the dataset may contain irrelevant information or artifacts from the background that could confuse the model. Removing unnecessary elements is crucial to ensure the model focuses only on the important features of the hand gesture. For instance, applying filters or segmentation techniques to separate the hand from the background can greatly improve the accuracy of recognition. Data preprocessing involves splitting the dataset into distinct subsets: training, validation, and testing. The training subset is used to teach the model, while the

validation subset helps optimize hyperparameters. The testing subset assesses the model's ability to generalize to new, unseen data. Proper preprocessing ensures that the data is organized, clean, and consistent, forming a solid foundation for building an effective machine learning model.

D. Segmenting the Gestures

A critical step in hand gesture recognition is isolating the hand from the background, a process known as segmentation. This step is essential as it directs the model's attention to the relevant part of the image—the hand—while disregarding the surrounding scene. However, segmentation can be challenging due to the variability in backgrounds across images and the potential for the hand to blend with different environments.

Several computer vision techniques are available for effective segmentation. Background subtraction is one such method, which works by comparing the current image frame to a background model and then identifying the regions that have changed. This method works best in controlled environments where the background remains constant. However, when the background is dynamic (e.g., a changing video stream), more advanced techniques such as skin color detection may be used. Skin color detection involves identifying pixels in the image that match the typical color range of human skin. This technique can help isolate the hand in situations where the background is complex. However, skin color detection can be sensitive to variations in lighting and skin tone, so it may need to be combined with other methods such as edge detection or depth sensing. Depth-based methods, which utilize stereo cameras or depth sensors, offer an alternative approach for segmenting the hand from the background. These methods work by emphasizing the disparity between the foreground (the hand) and the background, making it easier to distinguish the hand from its surroundings. These methods are particularly useful in environments with cluttered or visually complex backgrounds, as they rely on the three-dimensional properties of the scene. Additionally, techniques like semantic segmentation or instance segmentation can be utilized in machine learning. These methods employ neural networks to categorize each pixel in an image as either part of the hand or the background, enabling more precise and detailed segmentation. Once the hand is effectively isolated from the background, the next steps in processing— such as feature extraction and model training— can proceed more smoothly. Accurate segmentation ensures that the model concentrates on the correct object within the image, leading to a notable enhancement in its performance.

E. Transforming Gestures into Multiple Languages

A key strength of a hand gesture recognition system, especially one designed for American Sign Language (ASL), is its ability to not only convert gestures into text but also offer translations into various spoken languages. This capability is especially important for breaking down communication barriers in a multicultural and multilingual world. However, the task of transforming gestures into multiple languages is far from simple—it involves not only the recognition of the gesture itself but also understanding the meaning behind it and mapping it accurately to the corresponding text in the target language. The first step in transforming gestures into multiple languages is to recognize the gesture and translate it into a base language, usually English. This is the simplest form of translation, where each ASL gesture corresponds directly to an English letter or word. However, ASL is a complete language, with its own unique syntax and grammar, which may differ significantly from the structure of other languages. For instance, ASL follows a different word order than English and places significant emphasis on physical movements and facial cues to communicate meaning, aspects that cannot be directly translated into written words.

This makes a simple word-for-word translation system insufficient. To address this, a more sophisticated approach is required—one that understands the full context of the gesture. This includes considering the surrounding signs, the body posture, facial expressions, and hand movements. Within American Sign Language (ASL), the meaning of a single handshape can vary depending on the context. Therefore, the system must be capable of analyzing not only the gesture itself but also how it interacts with other gestures to construct a complete phrase or sentence. This is where Natural Language Processing (NLP) comes into play. By incorporating NLP models trained to understand both the syntax and semantics of ASL and its target language, the system can provide a more refined translation that accurately captures the meaning of the original sign. Once the gesture is recognized and understood in its context, the next step is to translate it into the target language. This involves

selecting the appropriate word or phrase in the target language, ensuring that the translation is both semantically and syntactically correct. For instance, the word "thank you" in ASL involves a gesture where the fingertips touch the chin and move outward, but the translation of "thank you" into other languages might involve different expressions or even different grammatical structures. In languages like Hindi the translation may follow a different word order or incorporate formal or informal forms of address, which must be handled by the translation system. One way to accomplish this is by using a machine translation (MT) system, which has been trained on parallel corpora—large datasets that contain the same text in multiple languages. These models use statistical methods, deep learning, or transformer architectures (such as Google's Transformer model) to map the structure of the source language (in this case, ASL) to the target language. However, machine translation is not always flawless and can lead to errors, particularly when translating between languages with significantly different linguistic structures.

To enhance translation quality, the system can incorporate postediting by human experts or feedback loops from users. For example, if the system makes an error in translating a specific sign or phrase, users can flag the issue, allowing the system to learn and improve from this feedback over time. This could be done via a user interface where the system provides an initial translation, and users can correct it if necessary, allowing the system to incorporate those corrections into its learning process. Another consideration when translating gestures into multiple languages is cultural context. Some signs in ASL might have different meanings or might not exist in other languages, requiring the system to either map them to the closest equivalent or to suggest alternative ways to express the same idea. For example, the ASL sign for "school" might not have a direct equivalent in other languages, or the gesture might be culturally specific. The system must take such cultural nuances into account, which may involve consulting additional data sources or databases that contain information about the cultural context of various languages and their gestures.

Real-time translation is another challenging aspect. Translating gestures into multiple languages must happen quickly to provide meaningful, interactive experiences for users. This involves fine-tuning the system to not only identify gestures accurately but also to convert them into text or speech quickly. For instance, when a user makes a sign, the system must process the gesture, interpret its meaning, and generate the translated output (whether in text or speech) with minimal delay. This realtime processing can be achieved through efficient model architectures, low-latency translation systems, and powerful hardware. In the context of multilingualism, the system should also be able to handle language switching, where the user may request translations in different languages based on their preference.

The system could offer a user interface where the user can select the language, they want the translation in or automatically detect the language preferences based on their location, profile, or previous interactions. Furthermore, by integrating voice recognition and synthesis systems, the system could offer spoken translations of gestures in multiple languages. This would provide a completely accessible experience for both people who are deaf or have hearing impairments and those who are learning the language. As the system evolves, it can also integrate advanced technologies like neural machine translation (NMT) and multilingual language models that learn from large, diverse datasets. This allows the system to better handle not only common words and phrases but also rare or complex expressions. For instance, some ASL signs may represent entire concepts or ideas that do not have a one-to-one equivalent in other languages. In these cases, the system must be capable of translating the meaning or concept behind the sign rather than just the word. Additionally, by accommodating local variations or forms of a language, (such as the differences between Spanish in Spain and Latin America), the system can become even more inclusive. By supporting a variety of dialects, the system can cater to a wider audience and increase the quality and relevance of the translations. Over time, the system can learn and adapt to the specific needs of different communities, improving its performance and usability.

F. Extracting Features

Extracting features plays a crucial role in the machine learning process for recognizing hand gestures [33]. The goal is to convert the raw data into a more meaningful format (such as the segmented image) into a set of meaningful features that the machine learning model can use to make accurate predictions [26]. For

hand gestures, these features often include both geometric and appearance-based characteristics.

Geometric features refer to the physical attributes of the hand and its pose. For example, factors such as the position of the fingers, the angles between the joints, and the overall shape of the hand serve as crucial indicators that help the model distinguish between various ASL gestures (Fig. 1). These features can be extracted by detecting key points or landmarks on the hand using algorithms like OpenPose or MediaPipe, which are capable of tracking the positions of various joints in real time. Appearance-based features refer to the texture, shape, and other visual aspects of the hand that help define its identity in the gesture. These features can be extracted using methods like Histograms of Oriented Gradients (HOG), which capture the contours and textures of the hand. Additionally, methods like Speeded-Up Robust Features (SURF) or Scale-Invariant Feature Transform (SIFT) can be utilized to capture distinctive patterns or features of the hand, which remain consistent despite changes in scale, rotation, or orientation. For dynamic gestures that involve movement over time, temporal features are essential.

These features track the changes in the hand's position and orientation across consecutive frames, enabling the system to recognize gestures involving motion or transitions. By combining both geometric and temporal features, the model gains a deeper understanding of the gesture, enhancing its ability to distinguish between similar signs. Feature extraction not only simplifies the data but also guarantees that the model concentrates on the most important details for precise predictions. Accurate feature extraction plays a key role in determining the performance of the gesture recognition system.
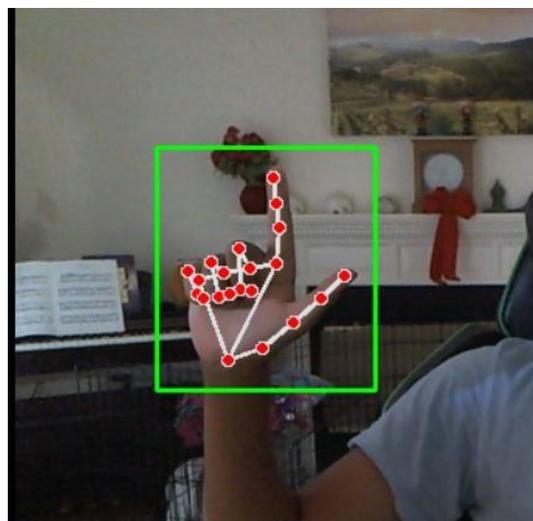

Fig. 1 Feature Recognition Using Coordinate Mapping

G. Training the Model

Once the dataset has been preprocessed, images segmented, and features extracted, the next step is to use the processed data to train a machine learning model that can reliably identify ASL gestures. One of the most commonly used and effective models for image-based tasks like this is the Convolutional Neural Network (CNN) [45]. Image data is processed by CNNs, which automatically detect patterns through a sequence of convolutional layers. The layers gradually extract hierarchical features from the raw pixel data, starting with fundamental elements like edges and textures, and advancing to more complex features like gestures, objects, and shapes.

The CNN can then classify the image based on these learned features which are extracted from datasets containing labeled examples, such as training data for specific ASL letters like B (Fig. 3). However, CNNs alone may not be sufficient to capture the full complexity of hand gestures, particularly for dynamic gestures that involve movement over time. In these cases, combining CNNs with Recurrent Neural Networks (RNNs) or Long Short-Term Memory (LSTM) networks offers a powerful solution. Integrating Recurrent Neural Networks (RNNs) with Long Short-Term Memory (LSTM) networks creates a strong solution that require understanding the temporal aspects of gestures. When combined with CNNs, RNNs or LSTMs can process both spatial features (such as the hand's shape and position) and temporal features (such as the hand's movement over time). This enables the model to comprehend gestures

that involve both static positions and dynamic transitions. Training the model involves feeding it a labeled dataset containing input images or video frames, along with the corresponding gestures or ASL letters as labels (Fig. 2).

The model uses these labels to adjust its internal parameters through a process called backpropagation [26]. During this process, the model computes the difference between its predicted output and the actual label, then adjusts its weights to reduce this error. This process is repeated over many iterations (epochs) until the model converges on an optimal solution. Since training can be computationally demanding, it often requires high-performance GPUs or cloud computing resources to accelerate the learning process. An important factor in training is choosing appropriate hyperparameters, including the learning rate, batch size, and the depth of the network. Methods for optimizing hyperparameters, such as grid search or random search, help in determining the most suitable set of parameters for the task.

To prevent overfitting, where the model performs well on the training dataset but struggles with new data, it is crucial to apply methods such as dropout or data enhancement [33]. After training the model, it is tested on a separate validation set to confirm its ability to generalize to unseen data. If the model demonstrates strong performance on the evaluation set, it can then be evaluated on a test set to estimate its real-world performance.

Fig. 2 Dataset used for training the model

Fig. 3 Training Data given for Letter B

H. Testing the Model

Testing plays a vital role in the development of a hand gesture recognition system as it evaluates the model's capacity to adapt to new, previously unseen data. During this stage, the model undergoes assessment with a distinct dataset that wasn't used during training, offering insights into its potential performance in real-world scenarios. Several metrics, such as precision, recall, accuracy, and F1 score, can be used to gauge the model's effectiveness. Although accuracy indicates the proportion of correct predictions, it may not fully reflect the model's performance, particularly in cases where the dataset is skewed (e.g., when some gestures are overrepresented). In situations with imbalanced datasets, precision and recall provide a deeper insight into the model's ability to accurately recognize various gestures. Precision measures the proportion of true positives out of all positive predictions, while recall evaluates the percentage of true positives among all actual positive instances. The F1 score, which merges both precision and recall, offers a comprehensive view of the model's performance, especially in cases where class distribution is uneven. These metrics ensure the model performs effectively across diverse gesture types, particularly in challenging or imbalanced scenarios.

Furthermore, confusion matrices are valuable tools for analyzing the model's errors. A confusion matrix shows the count of instances where the model correctly or incorrectly identified a specific gesture and compares it to the actual gesture, helping to identify where the model may be confusing certain gestures. By visualizing these discrepancies, the confusion matrix provides insights into which gestures are more likely to be misclassified, enabling focused enhancements in the model's effectiveness.

This data is crucial for recognizing trends in the model's errors and guiding adjustments. The testing phase also involves assessing the model's ability to perform in real-world settings, considering factors like changes in lighting, hand placement, and background elements. This ensures the model maintains its reliability and precision across various, unpredictable conditions. Since ASL gestures can be made with varying hand shapes, orientations, and speeds, it's important to evaluate how well the model can handle these variations. The testing dataset should ideally

cover these factors to ensure the model's robustness. In addition to accuracy, it is crucial to consider the speed and efficiency of the model, especially in scenarios that demand real-time processing, like live gesture detection. Ensuring that the model can process and interpret gestures quickly is crucial for delivering a smooth and uninterrupted interaction experience. In such cases, the model's prediction time (the duration it takes to produce an output after receiving input) should be minimized to ensure a fluid user interaction. If the model's performance during testing is not satisfactory, it may need further tuning, additional data, or enhancements to its architecture in order to improve its results. This iterative process helps refine the model and increase its accuracy and reliability [33].

I. Deploying the System

After the machine learning model undergoes training and evaluation, the next crucial step is deployment [33]. This entails incorporating the trained model into an operational system that can be accessed and used by end-users in real-world applications. For gesture recognition systems, deployment typically involves creating an application or software that utilizes a camera (such as a webcam or mobile phone camera) to capture the user's hand gestures, which are then sent to the trained model for recognition. The system should be able to handle video frames efficiently in real-time, requiring optimization to ensure the model works efficiently on the selected hardware. Deep learning models, particularly CNNs, can require significant computational resources, and running them in real-time on devices like smartphones or embedded systems demands reducing the model's size and computational load for smooth performance. Techniques like model quantization, pruning, or using more efficient architectures (e.g., MobileNet or EfficientNet) can be employed to make the model smaller and faster without sacrificing too much accuracy.

Another approach to real-time deployment is using frameworks like TensorFlow Lite or ONNX, which allow for the conversion of a trained model into a format optimized for mobile devices or embedded systems. In addition to model optimization, the success of the deployed system heavily depends on the design of the user interface (UI) and the overall user experience (UX). The application should be intuitive and easy to use, with clear instructions that guide users in interacting with the system effectively. A well-designed UI/UX ensures that users can navigate the system with ease, enhancing overall user satisfaction and engagement. For instance, it may include features like a live video feed of the user's hand, real-time feedback on detected gestures, and text output of the recognized ASL letters. Accessibility features like speech output or vibration alerts could also be integrated to enhance the system's inclusivity. Moreover, the application must be developed to accommodate various input conditions, such as fluctuations in lighting, background, and hand positioning. This guarantees the system's robustness and performs accurately in diverse environments, providing a seamless experience for users under different conditions.

Deploying a gesture recognition system also involves setting up a cloud infrastructure (if needed) to handle storage, processing, and serving predictions. In some cases, a cloud-based deployment may be preferred for models that require high computational power, as this would offload the heavy processing from the local device to the cloud, allowing the system to deliver accurate results even on less powerful devices. However, for real-time applications, deploying the model on the edge device itself (e.g., a smartphone or IoT device) is often more practical. Finally, once deployed, the system must be tested under real-world conditions to ensure that it performs reliably and accurately, even when faced with variations in user behavior or environmental factors.

J. Evaluating the Results

After deployment, the system must undergo continuous evaluation to ensure that it continues to function effectively and improve over time. One aspect of evaluation involves evaluating the model's effectiveness in real-world settings, where factors like hand positioning, lighting, and background interference can influence the accuracy of gesture recognition. Monitoring how the model adapts to these variables is crucial for ensuring its reliability and effectiveness in practical use. The real-world deployment might reveal issues that were not apparent during testing, such as the model struggling with certain gestures or performing poorly in particular environmental conditions. For example, the model may struggle to recognize gestures when the user's hand is too close to or too far from the camera, or when the lighting is dim or uneven. These elements can

degrade the quality of the input data, making it harder for the model to correctly recognize gestures [26]. To address these issues, the system can be designed to continuously collect user data (with consent) to identify areas where the model needs improvement. This data can be used to fine-tune the model and retrain it with more diverse and representative examples. Additionally, feedback from users, especially those with expertise in ASL, is invaluable for assessing the system's accuracy and usability. Such feedback helps identify areas for improvement, ensuring the system better meets the needs of its users. User feedback can highlight problems with the system, such as incorrect translations, missed gestures, or the need for additional signs to be recognized. Moreover, testing the system with a diverse group of users, representing various demographics such as age, gender, and ethnicity, can help identify potential biases in the model. This approach ensures that the system performs equitably and reliably for all users, regardless of their background or characteristics. Regularly collecting new data, retraining the model with updated datasets, and performing comparative testing with different iterations of the model can substantially boost the system's precision and enhance user experience. Furthermore, as the system evolves, new features, such as the ability to recognize dynamic gestures, entire phrases, or facial expressions, can be added to increase its functionality. Evaluation is a continuous process rather than a one-time activity, ensuring that the system stays effective and adapts to the evolving needs of its users.

This research methodology ensures a technically innovative and user-centric solution that bridges communication gaps while addressing the challenges faced by previous models. By offering real-time, accurate, and multilingual translations, this system marks a significant advancement in enhancing communication for the deaf and hard-of-hearing community, promoting increased access and equality.

IV. SYSTEM DESIGN AND IMPLEMENTATION

The proposed system for translating hand gestures into text and speech provides a robust solution to close communication gaps, particularly for individuals who depend on sign language for communication. The workflow of the system is depicted in Fig. 4, illustrating how gestures are captured, recognized,

translated, and outputted as text or speech. By employing advanced technologies in speech synthesis, language processing, and machine learning, the system allows for real-time, multilingual communication. Its design is meticulously crafted to ensure flexibility and modularity, allowing it to adapt to future advancements and integrate user feedback, ensuring continuous improvement and enhanced user experience. Key components include a gesture capture interface, a machine learning-based gesture recognition engine, a language translation module, and an output generation interface. Each of these components is crucial for the system's smooth operation, ensuring seamless translations from sign language to spoken language. By integrating these technologies, the system enables efficient communication for individuals with hearing impairments, improving their capacity to engage with the wider community.

An essential part of the system is the interface for capturing gestures, which uses cameras or specialized sensors, such as Leap Motion, to capture and capture hand movements in real-time. This interface guarantees accurate and timely detection of gestures, forming the foundation for the system's capability to understand and convert sign language. This interface ensures that hand movements are tracked with high accuracy and speed, enabling precise input for the recognition module. For this purpose, high-resolution cameras and depth sensors are employed to capture even the most subtle hand movements. The Leap Motion sensor is especially useful, as it utilizes infrared-based technology to track intricate hand motions with exceptional precision, ensuring accurate gesture identification and improving the system's capacity to understand intricate signs. It records precise movement data, making it an ideal option for gesture-driven applications. Additionally, the system's gesture capture module is optimized to support various hand sizes, shapes, and motions, guaranteeing flexibility for different users. It is also designed to function effectively in various lighting conditions, enhancing its robustness and reliability across different environments.
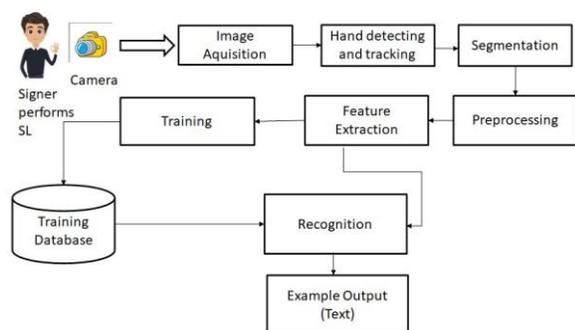
Fig. 4 Working of Sign Language Recognition System

Once the gestures are captured, the machine learning based recognition engine comes into play. This module employs advanced deep learning techniques to process the captured data and recognize the corresponding sign language gestures. Convolutional Neural Networks (CNNs) are highly effective for this task, as they are optimized for pattern recognition in visual inputs, making them particularly well-suited for detecting hand forms and actions from images or video frames. CNNs excel at extracting spatial hierarchies from images, enabling the system to accurately identify gestures. To train the recognition engine, a large, labeled dataset of hand gesture images is necessary, where each gesture is associated with a specific sign in the sign language being recognized. This dataset allows the CNN to learn the features that distinguish different gestures, improving its recognition capabilities. Over time, the system learns to identify key features of each gesture, such as hand shape, orientation, and movement. By leveraging large datasets, the recognition engine can continuously improve its accuracy and performance, adapting to various hand shapes, sizes, and environmental conditions [40]. This guarantees that the system stays consistent and efficient for different users, providing consistent gesture recognition across a range of scenarios.

The accuracy of the gesture recognition engine is a critical aspect of the system. A misinterpreted gesture can compromise the entire translation process, resulting in incorrect text or speech output. To overcome this challenge, methods such as transfer learning and data augmentation are utilized. Data augmentation creates additional training data by modifying the original dataset through techniques like rotation, scaling, and color variations. This strengthens the model's resilience, enabling it to better handle variations in hand gestures and boosting its performance in practical scenarios. Meanwhile, transfer learning enables the system to leverage pre-trained deep learning models for improved efficiency and accuracy, which have already learned to recognize key features from large datasets. This reduces the amount of training data needed, accelerates the training process, and ensures accuracy even with limited data. After the gesture is identified, the system moves on to translate it into the appropriate language, converting the identified gesture into English text as an intermediary step before translating it into the target language, such as Hindi.

The language translation module is responsible for converting the recognized English text into Hindi. This is accomplished through integration with the Google Translate API, a powerful translation service that provides fast and accurate translations. Google Translate leverages neural machine translation (NMT) models, which utilize deep learning techniques to surpass the limitations of traditional translation methods. These neural machine translation (NMT) models are trained on extensive bilingual datasets, allowing them to grasp the complex connections between words and phrases across languages, leading to more fluent and context-aware translations. With the integration of Google Translate, the system can deliver accurate and contextually relevant translations from English to Hindi.

Once the English text is translated into Hindi, the output generation interface takes over. This module is responsible for converting the translated Hindi text into both textual and speech output. The translated text is shown on the screen, allowing the user to easily read it, while the speech is generated using a text-to-speech (TTS) engine. This TTS engine translates the Hindi text into natural-sounding speech, ensuring that the translation is accessible both visually and audibly to users with different needs, enhancing the overall user experience. TTS technology has evolved significantly over the years, with many modern systems, such as Google Cloud Text-to-Speech and others, offering highly natural-sounding speech in various languages, including Hindi. These systems are trained on large speech datasets to produce speech that closely mimics human conversation. The TTS engine in the system allows users to hear the translated message in Hindi, enhancing the accessibility for individuals who may prefer auditory communication or those with visual

impairments. The hardware requirements for the system are essential for ensuring its high performance. The gesture capture module requires a high-resolution camera or specialized to capture detailed hand gestures. The machine learning recognition engine demands significant computational power, particularly when using deep learning models. To achieve real-time processing, the system must be deployed on a server equipped with a Graphics Processing Unit (GPU). The GPU accelerates the processing of complex deep learning models, enabling faster computations and efficient handling of large volumes of data, which is essential for tasks like gesture recognition and language translation in real time.

A GPU accelerates the parallel processing needed for deep learning algorithm allowing the system to process gestures and generate translations quickly. Additionally, the system requires a display to show the text output and a speaker to deliver the speech output. Optional hardware, such as haptic feedback devices, could be integrated to further enhance the user experience, particularly for individuals with visual impairments. These devices would provide tactile feedback when gestures are recognized, offering an additional layer of accessibility. The software implementation of the system primarily relies on Python, a language commonly used for machine learning and AI applications. Libraries like TensorFlow and PyTorch are leveraged to develop the gesture recognition model, as they provide robust tools for constructing and training deep learning networks, enabling efficient and accurate gesture detection and translation.

These frameworks allow for efficient model training and fine-tuning, ensuring the recognition engine is capable of accurately identifying gestures. For language translation, the system integrates Google Translate API, while speech generation is handled by Google Cloud Text-to Speech or similar services. The entire system is designed in a modular format, ensuring ease of future updates and expansions. As new languages, gestures, or machine learning models become available, they can be smoothly incorporated into the system without affecting its overall operation. The system undergoes iterative testing to evaluate its accuracy, speed, and usability. The gesture recognition engine, translation module, and text-to-speech (TTS) system undergoes extensive testing in diverse scenarios to validate its performance. Key performance indicators, including efficiency, response time, and recognition accuracy, are evaluated under real-world conditions. Furthermore, direct input from users is collected to uncover any limitations or areas requiring enhancement. This iterative process of evaluation and improvement ensures the system delivers consistent and reliable outputs, accurately converting gestures into text and spoken language. By bridging communication gaps for sign language users, the developed system marks a meaningful step forward in facilitating inclusivity and accessibility. By leveraging machine learning, the Google Translate API, and text-to-speech technologies, the system offers an accessible and effective communication platform for diverse users.

The modular architecture allows the system to evolve with future advancements in technology, ensuring its long-term relevance and utility. The focus on Hindi as a target language expands the system's applicability to a large user base, particularly in regions where Hindi is widely spoken, such as in India. The system's flexibility, real-time performance, and seamless integration of gesture recognition, language translation, and speech synthesis make it an essential resource for enhancing communication and fostering inclusivity for people with hearing impairments.

V. PROBLEM STATEMENT

The divide between the hearing and deaf populations continues to be a critical societal concern, significantly hindering the participation of individuals with hearing impairments in various domains, such as professional, cultural, and social activities, as well as access to education. As a result, the deaf community often experiences feelings of isolation and exclusion, unable to fully integrate into a world that largely revolves around spoken communication. While many countries have taken steps to bridge this divide through the development of sign language interpretation systems, the existing solutions continue to face significant hurdles that limit their accessibility, accuracy, and inclusivity. These issues are particularly pronounced in diverse and multi-lingual societies like India, where the existing technological infrastructure and sign language models fail to address the full spectrum of sign languages used by different communities.

Drawbacks of Conventional Sign Language Translation

Traditional methods of sign language interpretation, whether through human interpreters or video-based solutions, are inherently limited in their ability to scale and to effectively address the diverse needs of the deaf population, solutions must be adaptable and inclusive. The main challenge with traditional human interpreters is their limited availability, restricting deaf individuals' access to real-time interpretation when it is most crucial. This limitation becomes more apparent in emergency situations, public spaces, or areas with limited access to professional interpreters. Furthermore, the availability of skilled interpreters is uneven, especially in rural or remote areas, leaving many individuals without the support they need to engage with the hearing world.

Sign language recognition systems that rely on video for automated interpretation face challenges in technology and accuracy. These systems capture user gestures via video and convert them into text or speech. However, varying lighting conditions can significantly impact the system's recognition accuracy. Poor lighting, shadows, or bright light sources can obscure hand movements and reduce recognition accuracy. Similarly, variations in camera resolution or the presence of background disturbances can interfere with the system's ability to capture clear, high-quality data.

Additionally, these systems struggle with variations in gesture vocabulary. Different sign languages, even within the same country, may have unique signs and variations of the same signs that cannot always be standardized in a way that ensures universal recognition. A system built to recognize American Sign Language (ASL) may fail to interpret Indian Sign Language (ISL) due to the distinct cultural, regional, and contextual characteristics of each sign language. For this reason, gesture-based systems must account for a wide variety of regional dialects, hand shapes, movements, and individual signing styles, which further complicates the challenge.

Specific Challenges in India

In India, the issue of sign language communication is compounded by the country's complex linguistic landscape. India hosts more than 22 constitutionally acknowledged languages and a vast array of regional vernaculars. This linguistic diversity makes it difficult for a single sign language system to cater to the needs of all communities. Indian Sign Language (ISL), for instance, differs not only from other sign languages like ASL but also contains region specific variations that make it even more challenging to create a uniform system for all users. In urban areas like Delhi or Mumbai, a relatively standardized form of ISL might be used, but in rural areas, local variations of ISL or even entirely different regional sign languages may be prevalent.

The lack of multilingual support in current gesture recognition systems further exacerbates this issue. Most systems are designed with a primary focus on English or a small set of widely spoken languages, with little consideration given to the diverse linguistic needs of countries like India. This leaves individuals who rely on ISL or other regional sign languages in a precarious position, unable to access meaningful communication in their native languages. The failure to incorporate regional sign languages into gesture recognition systems in a way that respects the nuances of these languages leads to misunderstandings, misinterpretations, and, ultimately, social exclusion.

Moreover, the absence of effective solutions for multilingual sign language translation is a significant barrier to full social integration. Deaf individuals in India who communicate using ISL or other regional sign languages often find themselves excluded from public services, education, healthcare, and other essential areas of life. This exclusion can have serious long-term consequences, hindering their ability to access resources, build meaningful social connections, and engage in career development opportunities. In a country where language barriers are already a significant issue for many people, the deaf community is doubly marginalized, with little recourse for addressing these challenges through technology.

Technological Barriers

Despite advancements in gesture recognition systems driven by machine learning and artificial intelligence, these technologies remain underutilized in meeting the specific needs of the deaf community. Many current systems fall short of delivering the accuracy needed for real-time sign language interpretation, especially in dynamic and unpredictable settings. The models used in these systems must be able to recognize gestures in varying conditions, such as changing lighting, noise, or even when users are wearing different clothing that may obscure hand gestures.

The lack of standardization in training datasets is another major barrier to the development of more inclusive systems. High-quality datasets that represent

the full spectrum of hand gestures, including regional variations of sign languages, are often scarce or inaccessible. This gap in data availability means that machine learning models trained on limited datasets cannot generalize well to unseen gestures or users, resulting in poor recognition performance. Moreover, the computational requirements for training accurate models are high, making it difficult for developers to create systems that are both accurate and available to users with minimal technological capabilities.

The challenge of recognizing gestures in real-time is further intensified by the requirement for advanced language translation. While machine learning models can convert sign language gestures into text, translating that text into regional languages introduces additional intricacy. For instance, Hindi, a predominant language spoken by millions across India, has its own set of syntactic and grammatical rules that differ significantly from English. This distinction creates challenges for translation, as direct word-for-word translation can lead to awkward phrasing or lost nuances. This is particularly crucial in sign language systems, where signs are often context-dependent and require more than a basic translation.

The Need for a User-Centric Design

Another fundamental problem with current gesture recognition systems is their lack of user centric design. Many of the existing systems are highly technical and require users to have a certain level of familiarity with technology to operate effectively. For the deaf community, this can present a significant barrier. Many deaf individuals may not be technologically savvy or may face additional challenges related to literacy or familiarity with computers and smartphones. For these systems to be truly inclusive, they should be crafted with a focus on the needs and preferences of the end-user, prioritizing ease of use and accessibility over complexity.

Apart from ease of use, it is essential for these systems to be adaptable and function effectively across various settings and circumstances. The deaf community often faces barriers to communication in public spaces, hospitals, and schools, where access to interpreters or assistive technologies is limited.

A system capable of operating in such diverse settings without requiring expensive or specialized hardware would greatly enhance the ability of deaf individuals to communicate in everyday situations. This requires the system to be lightweight, scalable,

and robust enough to handle various external factors like changes in lighting, noise, and user movement.

Proposed Solution

The objective of this study is to overcome these obstacles by developing a comprehensive sign language translation solution, integrating advanced machine learning techniques, particularly emphasizing convolutional neural networks (CNNs) for accurate gesture recognition. The proposed system aims to translate hand gestures into English text, followed by a module that converts this text into Hindi, India's most widely spoken language. By leveraging deep learning to adapt to regional sign languages, the system will enable users to communicate using their native signs, irrespective of geographic location or dialect.

In addition to addressing the technical limitations of existing systems, the proposed solution will prioritize accessibility and ease of use. The system will be designed to function efficiently in dynamic environments, ensuring that it has the potential to be applied in various environments, such as public spaces, low-light areas, and crowded environments. The system will also be optimized for various hardware, including smartphones, laptops, and specialized gesture capture devices like the Leap Motion sensor, ensuring accessibility for individuals with limited access to advanced equipment.

The objective of this research is to design a system capable of identifying different gestures, translates them into multiple languages, and provides real-time feedback, ultimately enhancing communication between deaf and hearing individuals. By integrating machine learning algorithms, multilingual translation, and user-centered design, the system seeks to promote a society that is more inclusive and accessible by removing obstacles to communication that contribute to isolation and exclusion.

The central focus of this study is to improve societal inclusion and communication for all individuals and opportunities for the deaf community, equipping them with the tools and resources necessary to succeed in a digitally connected world.

VI. CONCLUSION

This research represents significant progress in sign language interpretation by proposing A system designed to seamlessly close communication barriers

for the hearing-impaired community. By incorporating cutting-edge technologies such as deep learning, gesture detection, and multilingual translation, the system provides a holistic approach to addressing the shortcomings of earlier models.

By integrating features such as real-time processing, support for the most widely spoken language in India, and an intuitive user interface, the system ensures accessibility and inclusivity for a diverse user base. The modular design improves adaptability and scalability, making the system versatile for different applications and environments. Comparative evaluations and user feedback affirm the system's improved accuracy, efficiency, and usability compared to earlier approaches. Future research may focus on further expanding gesture vocabularies, refining processing speeds, and exploring hybrid systems that combine camera-based recognition with wearable technology. By focusing on user-centric design and technological innovation, the suggested framework establishes a strong base for further progress in supportive communication systems.

This work represents a vital step toward empowering individuals with hearing impairments, fostering greater inclusivity, and enhancing accessibility in both domestic and worldwide contexts.

REFERENCES

[1] K. L. Decker and S. M. Hardesty, "The current rural sign language interpreter landscape: Examining the field's collective accountability," International Journal of Interpreter Education, vol. 15, no. 1, p. 11, 2024.

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

[3] Y. Gu, H. Oku, and M. Todoh, "American sign language recognition and translation using perception neuron wearable inertial motion capture system," Sensors, vol. 24, no. 2, p. 453, 2024.

[4] N. K. Bahia and R. Rani, "Multi-level taxonomy review for sign language recognition: Emphasis on Indian sign language," ACM Transactions on Asian and Low-Resource Language Information Processing, vol. 22, no. 1, pp. 1–39, 2023.

[5] H. ZainEldin et al., "Silent no more: A comprehensive review of artificial intelligence, deep learning, and machine learning in facilitating deaf and mute communication," Artificial Intelligence Review, vol. 57, no. 7, p. 188, 2024.

[6] J. M. De Martino et al., "Neural machine translation from text to sign language," Universal Access in the Information Society, pp. 1–14, 2023.

[7] T. Ashby, B. K. Webb, G. Knapp, J. Searle, and N. Fulda, "Personalized quest and dialogue generation in role-playing games: A knowledge graph-and language model-based approach," in Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, pp. 1–20, 2023.

[8] Y. Li, H. Wang, Q. Jin, J. Hu, P. Chemerys, Y. Fu, Y. Wang, S. Tulyakov, and J. Ren, "Snap fusion: Text-to-image diffusion model on mobile devices within two seconds," in Advances in Neural Information Processing Systems, vol. 36, 2024.

[9] H. Qu, X. Hui, Y. Cai, and J. Liu, "LMC: Large model collaboration with cross-assessment for training-free open-set object recognition," in Advances in Neural Information Processing Systems, vol. 36, 2024.

[10] T. Gupta and A. Kembhavi, "Visual programming: Compositional visual reasoning without training," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 14953–14962, 2023.

[11] B. Fu, P. Ye, L. Zhang, P. Yu, C. Hu, X. Shi, and Y. Chen, "A token-level contrastive framework for sign language translation," in ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1–5, 2023.

[12] C. H. Song, J. Wu, C. Washington, B. M. Sadler, W.-L. Chao, and Y. Su, "LLM-planner: Few-shot grounded planning for embodied agents with large language models," in Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 2998–3009, 2023.

[13] J. Singh, K. Devkar, G. Manjwani, A. Chandra, and S. P. Bholane, "Sign language translator using deep learning," International Journal of Advanced Research in Science, Communication and Technology, pp. 482–487, 2022. [Online]. Available: https://doi.org/10.48175/ijarsct-3690

[14] K. R. Prabha, M. Mahalingamuthuram, C. S. Mukesh, and V. P. P. Dharshan, "IoT-enabled sign language interpreter glove," pp. 601–606, 2024. [Online]. Available: https://doi.org/10.1109/icosec61587.2024.10722117

[15] H. Alzarooni, H. Al-Ali, S. Fadaaq, and T. Bonny, "AI-based sign language interpreter (GestPret)," pp. 79–85, 2023. [Online]. Available: https://doi.org/10.1145/3633598.3633612

[16] H. Qu, Y. Cai, and J. Liu, "LLMs are good action recognizers," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024.

[17] H. Liu, C. Li, Q. Wu, and Y. J. Lee, "Visual instruction tuning," in Advances in Neural Information Processing Systems, vol. 36, 2024.

[18] M. Gawande, S. Haral, P. Maknor, and V. Nerpgar, "Sign Language Recognition," IJIRMPS, vol. 12, no. 6, pp. --, Nov.-Dec. 2024. Paper ID: 231556. Published on: Nov. 11, 2024.

[19] S. M. Miah, M. A. M. Hasan, S. W. Jang, H. S. Lee, and J. Shin, "Multi-stream general and graph-based deep neural networks for skeleton-based sign language recognition," Electronics, vol. 12, no. 13, p. 2841, 2023.

[20] A. Yin, T. Zhong, L. Tang, W. Jin, T. Jin, and Z. Zhao, "Gloss attention for gloss-free sign language translation," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2551–2562, 2023.

[21] A. S. Nikam and A. G. Ambekar, "Sign Language Recognition Using Image-Based Hand Gesture Recognition Techniques," in Online International Conference on Green Engineering and Technologies, 2016, doi: 10.1109/GET.2016.7916786.

[22] S. Daniels, N. Suciati, and C. Fatichah, "Indonesian Sign Language Recognition using YOLO Method," IOP Conference Series: Materials Science and Engineering, vol. 1077, p. 012029, 2021, doi: 10.1088/1757-899X/1077/1/012029.

[23] Faria, B.M., et al., Evaluation of distinct input methods of an intelligent wheelchair in simulated and real environments: a performance and usability study. Assistive technology: the official journal of RESNA, 2013. 25(2): p. 88-98

[24] Blum, A. and T. Mitchell, Combining labeled and unlabeled data with co-training, in Proceedings of the eleventh annual conference on Computational learning theory1998, ACM: Madison, Wisconsin, United States. p. 92-100.

[25] Zhang, D. and G. Lu, A Comparative Study on Shape Retrieval Using Fourier Descriptors with Different Shape Signatures. Journal of Visual Communication and Image Representation, 2003(14 (1)): p.41-60.

[26] V. S. Sharma, S. Mahajan, A. Nayyar, and A. K. Pandit, Deep Learning in Engineering, Energy, and Finance: Principles and Applications. 1st ed. Boca Raton, FL: CRC Press, 2024, ISBN 9781032627724.

[27] Andersen, M.R., et al., Kinect Depth Sensor Evaluation for Computer Vision Applications, in technical report ECE-TR-6 2012, Department of Engineering – Electrical and Computer Engineering, Aarhus University. p.37.

[28] Kauppinen, H., T. Seppanen, and M. Pietikainen, An experimental comparison of autoregressive and Fourier-based descriptors in 2D shape classification. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1995. 17(2): p. 201-207.

[29] Zafrulla, Z., et al., American sign language recognition with the kinect, in 13th International Conference on Multimodal Interfaces2011, ACM: Alicante, Spain. p.279-286.

[30] Andersen, M.R., et al., Kinect Depth Sensor Evaluation for Computer Vision Applications, in technical report ECE-TR-6 2012, Department of Engineering – Electrical and Computer Engineering,Aarhus University. p.37.

[31] Madankar, M., Chandak, M. B., & Chavhan, N. (2016). Information retrieval system and machine translation: A review. Procedia Computer Science, 78, 845-850. https://doi.org/10.1016/j.procs.2016.02.071

[32] Lieberman, Z., T. Watson, and A. Castro. Open Frameworks. 2004 10 October 2013 [cited 2011; open Frameworks is an open-source C++ toolkit designed to assist the creative process by providing a simple and intuitive framework for experimentation. Available from: http://www.openframeworks.cc/

[33] V. Sharmila, S. Kannadhasan, A. R. Kannan, P. Sivakumar, and V. Vennila, Challenges in

Information, Communication, and Computing Technology: Proceedings of the 2nd International Conference on Challenges in Information, Communication, and Computing Technology (ICCICCT 2024), April 26th & 27th, 2024, Namakkal, Tamil Nadu, India. 1st ed. Boca Raton, FL: CRC Press, 2024, ISBN 9781032906607.

[34] F. Lu, F. Tian, Y. Jiang, X. Cao, W. Luo, G. Li, X. Zhang, G. Dai, and H. Wang, "Shadowstory: creative and collaborative digital storytelling inspired by cultural heritage," in Proceedings of the the SIGCHI Conference on human factors in computing systems. ACM, 2011, pp. 1919–1928.

[35] S. Mystakidis, N. Lambropoulos, H. M. Fardoun, and D. M. Alghazzawi, "Playful blended digital storytelling in 3d immersive e-learning environments: a cost-effective early literacy motivation method," in Proceedings of the 2014 workshop on interaction design in educational environments. ACM, 2014.

[36]] J. Rubart, "Face-to-face collaboration points in storytelling using multitouch tabletop systems," in Proceedings of the 2015 workshop on narrative and hypertext. ACM, 2015, pp. 33–35.

[37] H. Liang, J. Chang, I. K. Kazmi, J. J. Jian, and P. Zhang, "Hand gesture-based interactive puppetry system to assist storytelling for children," Vis Comput, vol. 33, no. 4, pp. 517–531, 2017.

[38]] E. Ong, H. Cooper, N. Pugeault, and R. Bowden, "Sign language recognition using sequential pattern trees," in Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Providence, RI. IEEE, 2012, pp. 2200–2207.

[39] S. Sako and T. Kitamura, "Subunit modeling for japanese sign language recognition based on phonetically depend multi-stream hidden markov models," Universal Access in Human-Computer Interaction, pp. 548–555, 2013.

[40] Ghayoumi, M. (2023). Generative Adversarial Networks in Practice. CRC Press.

[41] Gururaj, H. L., Flammini, F., Srividhya, S., Chayadevi, M. L., & Selvam, S. (2024). Advances in Manufacturing IV. Springer Science and Business Media LLC.

[42] Mariprasath, T., Reddy Cheepati, K., & Rivera, M. (2023). Practical Guide to Machine Learning. CRC Press.

[43] Thahseen, Ahamed & Tissera, Wishalya & Vidhanaarachchi, Samitha & Aaron, N & Rajapaksha, Damitu & Fernando, P.V & Dias, Thisuru & Fernando, Thatamathy. (2023). Smart System to Support Hearing Impaired Students in Tamil. 1-8. 10.1109/SCSE59836.2023.10215039.

[44] Chaowanawatee, Kullawat & Silanon, Kittasil & Kliangsuwan, Thitinan. (2023). Thai Finger-Spelling using Vision Transformer. International Journal of Advanced Computer Science and Applications. 14. 10.14569/IJACSA.2023.0141135.

[45] A. Dagur, K. Singh, P. S. Mehra, and D. K. Shukla, Artificial Intelligence, Blockchain, Computing, and Security, Volume 2: Proceedings of the International Conference on Artificial Intelligence, Blockchain, Computing, and Security (ICABCS 2023), Gr. Noida, UP, India, 24-25 February 2023. 1st ed. Boca Raton, FL: CRC Press, 2023, ISBN 9781032678412.