Multi-Modal Sign Language Detection: Integrating Sign, Text, Image, and Voice

Arushi Gupta
KIET Group of Institutions
Ghaziabad, Uttar Pradesh
arushi.gupta@kiet.edu

Shitiz Rajvanshi
KIET Group of Institutions
Ghaziabad, Uttar Pradesh
shitiz.2024cs1143@kiet.edu

Taniya Singh
KIET Group of Institutions
Ghaziabad, Uttar Pradesh
taniya.2024cs1094@kiet.edu

Shubham Goel
KIET Group of Institutions
Ghaziabad, Uttar Pradesh
shubham.2024cs1035@kiet.edu

ABSTRACT- With recent advancements in a variety of methodologies, the field of study on sign language recognition is expanding quickly. The goal of this research is to create a system that is easy to use for people who have trouble speaking and hearing, especially those who use sign language. Sign language is very crucial for people who are vocally or audible impaired. For these people this is the only way of communication. Our project focuses on easing the communication process for those people. The primary objective of our project is to develop an application that will convert sign language (signs) as input into text and voice output and vice versa. The secondary objective of our application is to employ these features using an android application that can be used easily and should have an interactive UI that in turn enhances the overall experience of using the application.

Keywords– Text, Sign-recognition, Sign Language, ML-kit, Google Cloud Vision API, Android, Firebase, CNN

1) INTRODUCTION

The capabilities of Android apps have been greatly improved, enabling Java programs to run on mobile devices. Thanks to this advancement, people all over the world can now use their mobile devices to read and write emails, browse websites, and play Java games. Taking note of this development, we suggest using Android applications to improve communication. The introduction of SMS and MMS made it easier for deaf people, who had not often used cell phones, to communicate remotely. Deaf people can now communicate with both hearing and deaf people via texts. Even though there are dumb or deaf people all around us, many people find it difficult to communicate with them. There is a need for a solution that makes communication easier for everyone because avoiding interaction is not a solution. We have created an application to address this demand and facilitate users' everyday communication. Even as technology develops further, its application should constantly aim for advantages. Our program tries to make it easier for those who are dumb or deaf to communicate with others around them. While other developers have tried to improve sign language apps, our goal is to make ours more dependable and effective. The focus of current sign language apps is typically on text-to- sign or sign-to-text conversion. On the other hand, our program consists of two modules: Sign to Text and Text to Sign. Furthermore, our application enables users to upload their own images, cropping them after taking a picture or choosing one from the gallery. Next, the image's text is shown on the screen, and a sign language version of it is produced. audio to Sign Conversion, which converts audio memos or talks to text and then back to sign language, is another noteworthy feature. Because it eliminates the need for typing, this feature is very helpful for English speakers who need to translate text into sign language for greater understanding. Among the numerous difficulties confronted by the deaf and dumb is object recognition. With the help of our software, users may identify objects in an image by taking a picture or choosing one from the gallery,

without having to depend on others to identify them. The app also shows the percentage by which the image matches the object. Language recognition is our app's last functionality. When users enter text in a language they are not acquainted with, the app will translate it into American English, recognize it, and create the appropriate gesture image.

2) LITRATURE REVIEW

In [1] Sign Language especially Indian Sign Language (ISL), for the deaf and mute. It notes limited research post-ISL standardization, focusing on static hand gestures with minimal attention to dynamics. Despite efforts on ISL alphabet recognition, the process involves multiple stages, surveyed to assess research progress. In [2] an intelligent system for translating sign language to text, comprising hardware and software components. The hardware incorporates flex, contact, and inertial sensors on a glove. Software features a classification algorithm leveraging k nearest neighbours, decision trees, and dynamic time warping, enabling static and dynamic gesture recognition. In [3] three methods for sub-unit-based sign recognition. Boosting is employed to learn appearance-based sub-units, merged with a second-stage classifier for word-level sign learning. Another approach integrates 2D tracking-based sub-units with appearance-based handshape classifiers. The final method translates these into 3D, enabling real-time, user independent recognition of isolated signs. In [4] a deep convolutional neural network for direct classification of hand gestures in images, eliminating the need for segmentation or detection stages. In [5] two novel hand gesture recognition approaches for real-time sign language comprehension. Employing a hybrid feature descriptor merging SURF and Hu Moment Invariant methods yields a strong recognition rate. SURF and moment invariant features exhibit resilience to diverse variations, ensuring effective real-time performance. [5] Introduces two innovative methods for real-time recognition of hand gestures in sign language. These methods merge SURF and Hu Moment Invariant techniques into a combined feature descriptor, improving recognition accuracy while maintaining low time complexity. They also introduce derived features and utilize KNN, SVM, and HMM for classification, demonstrating enhanced real-time efficiency and robustness. [6] Presents novel strategies for real-time recognition, translation, and video production in Sign Language (SL). Employing MediaPipe and hybrid CNN + BiLSTM models for recognition, and NMT + GAN models for video generation, achieving classification accuracy exceeding 95%. Evaluation metrics reveal substantial enhancements, including a 38.06 BLEU score and impressive visual quality. [7] Addresses the challenges of Continuous Sign Language Recognition (CSLR) by introducing SignBERT, a deep learning framework merging BERT and ResNet. Outperforming conventional methods in accuracy and word error rate on demanding datasets, SignBERT underscores its effectiveness in modeling sign languages and extracting spatial features for real-time CSLR. [8] Examines sign language research, particularly vision-based hand gesture recognition systems from 2014 to 2020. Through analysis of 96 articles, it identifies key research areas: data acquisition, environment, and gesture representation. Signer-dependent recognition averages 88.8%, while signer independent recognition averages 78.2%, indicating opportunities for improvement, especially in continuous gesture recognition. [9] Introduces a dynamic hand gesture recognition system leveraging multiple deep learning architectures. Evaluated on a challenging dataset, it outperforms existing methods, demonstrating effectiveness in uncontrolled environments with diverse gestures. [10] Presents a real-time hand gesture recognition system utilizing a cost-effective webcam and image processing techniques. The system comprises four stages: image preprocessing, region extraction, feature extraction, and matching, achieving a 90.19% recognition rate for American Sign Language (ASL) alphabet gestures under various lighting and hand conditions.

3) SYSTEM REQUIREMENTS

A) HARDWARE REQUIREMENTS

Specific hardware components for the mobile application are necessary for the system's overall successful operation. To support the system's various functions, an Android smartphone running Android version 5.0 or higher is required. The gadget ought to possess a rear camera that can recognize motions in American Sign Language (ASL), enabling the visual input necessary for efficient communication. A microphone is also necessary to detect human voice.

B) SOFTWARE REQUIREMENTS

1. **Android Studio:** The official integrated development environment (IDE) for the Android operating system from Google is called Android Studio. Specifically designed for Android programming, this IDE is built on JetBrains' IntelliJ IDEA software.
2. **Google Vision API:** This API is a component of the Google API family and offers application programming interfaces (APIs) for integrating with other services and communicating with a range of Google

services. Google Maps, Gmail, Search, and Translate are a few examples. These APIs can be used by third-party apps to expand or improve the features of already-available services. Specifically, the Google Vision API provides features including analytics, user data access, and machine learning as a service (the Prediction API).

3. Firebase ML Kit: This mobile software development kit (SDK) makes use of Google's machine learning know-how to improve applications for iOS and Android devices. For developers with varying levels of experience, it provides a robust and intuitive package for integrating machine learning features. It only takes a few lines of code to add machine learning capabilities to an app with ML Kit; developers no longer need to be experts in neural networks or model optimization.

4. HashMap Class: The HashMap function makes it easier to map processed input to the database that is stored. The Map interface is implemented by the HashMap class, which enables the storing of key-value pairs where the keys need to be distinct. This class is part of the java.util package.

4) FLOWCHARTS AND DIAGRAMS

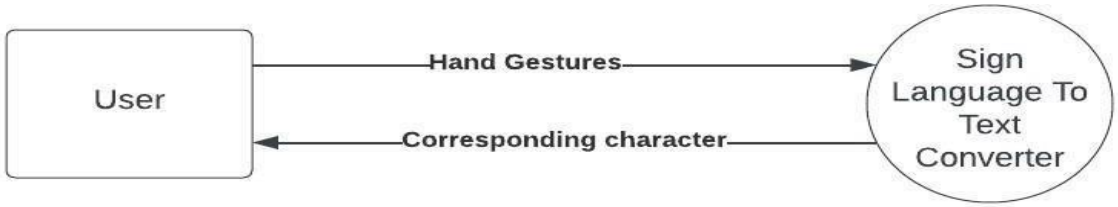

Figure 1: DFD Level-0 Diagram

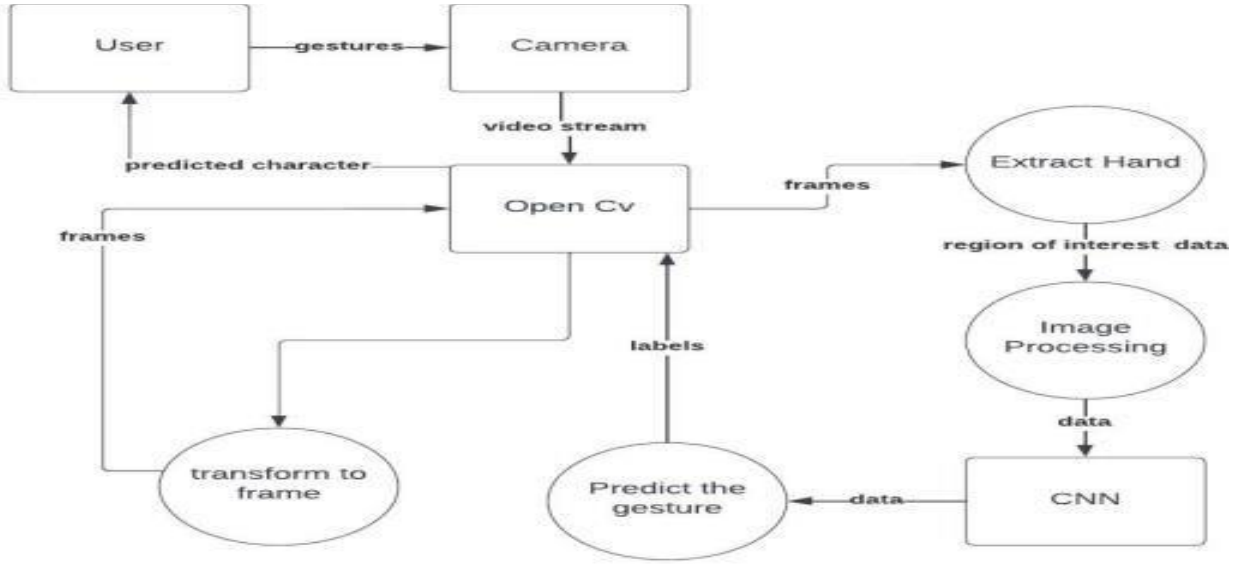

Figure 2: DFD Level-1 Diagram

\

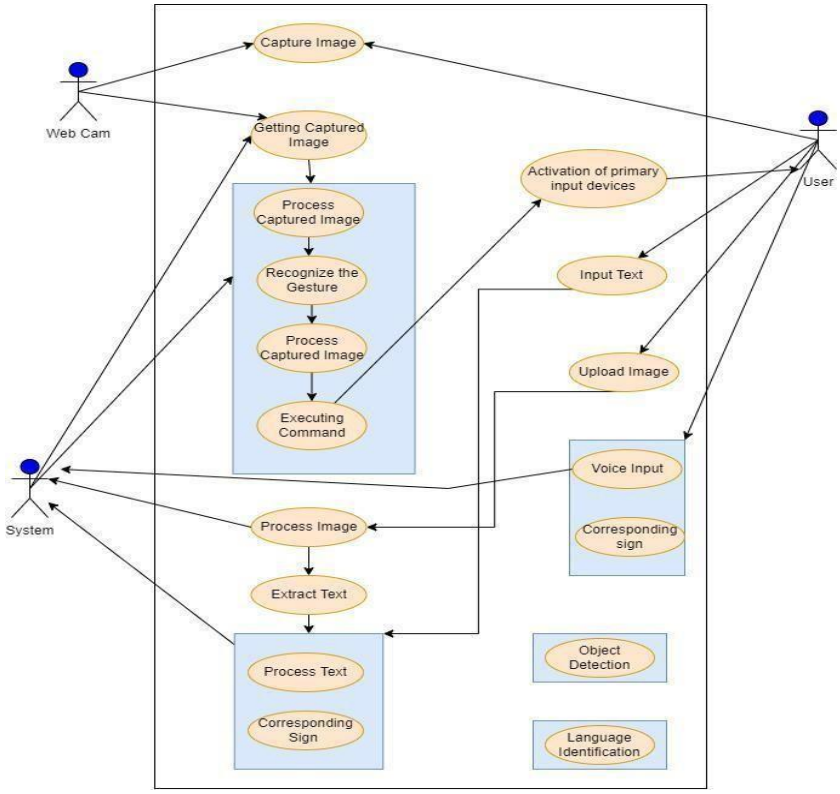

Figure 3: Use Case Diagram

5) Proposed System

For those who are deaf or mute, this app offers a ground-breaking alternative that makes communication easier. in the context of their environment. voice-to-sign transmission, vehicle voice recognition, and automatic translation are all supported by the proposed technology. In contrast to the current systems that mostly concentrate on one- way communication, our proposal offers the ability to communicate in both directions, from sign language to text or the other way around. Users can interpret texts into sign language and texts into sign language, which promotes better communication and educational opportunities. We will use Support Vector Machines (SVM) for both regression and classification techniques in our research. This method will be used to classify signs to determine how well they are functioning. A rich image dataset is essential for the implementation of sign language. At least 20 distinct representations of each alphabet should be used, considering changes in background, lighting (bulb, lamp, camera flash, daylight, and dim light), and distance. Our attention turns to improving image recognition for effective alphabet identification after the dataset is established. Text recognition will improve if conditions for text- to-image matching are included. Using a text-to-sign language interpreter, the text can be recognized and then shown as an image or animation in sign language. Whether an image is taken with the camera or imported from the gallery, every element in the picture will be examined in percentage terms to provide specific details. We will include a section that translates spoken languages from many nations, supporting as many languages as we can. This feature facilitates seamless communications by enabling users to comprehend the language spoken by others and ensuring worldwide accessibility. For thorough communication, each language type will be translated into a coherent English sentence and then into sign language.

6) System Design

Framework Plan is primarily partitioned into six segments: 1. Text to Sign Change

2. Picture to Sign Transformation

3. Voice to Sign Change

4. Sign to Text Transformation

5. Object Detection

6. Language Identification

These segments perform their exercises utilizing three important modules:

A. Content Acknowledgment Framework

B. Firebase Machine Learning Unit

C. Motion Acknowledgment Framework

The Framework Plan may be a comprehensive system that envelops six unmistakable segments, each serving as a specialized aspect of the system's capabilities. These segments incorporate Text to Sign Change, Picture to Sign Conversion, Voice to Sign Transformation, Sign to Text Change, Object Identification, and Language Identification. Together, they frame a cohesive design that addresses different modes of interaction and acknowledgment.

Central to this plan are three basic modules that collaborate to guarantee exact and seamless changes. The Content Acknowledgment
Framework, the primary module, serves as the spine of the system's capacity to interpret content input into expressive sign dialect signals. Firebase Machine Learning Unit, the moment centre component, essentially improves acknowledgment capabilities over a run of input modes, improving the system's precision and unwavering quality. The Motion Acknowledgment Framework, the third module, plays an essential part in translating perplexing hand developments and motions, contributing to both content and sign dialect transformations.

Text to Sign Transformation, the system's introductory area, enables the interpretation of written literary input into meaningful visual sign dialect expressions. This handle guarantees that clients can easily pass on their messages in a way that's all around caught on, bridging the gap between written dialect and the nuanced dialect of signs. Picture to Sign Change takes after, utilizing progressed algorithms to handle pictures and create comparing sign dialect motions. This include is especially useful when managing with visual substance, enhancing communication through visual prompts. Voice to Sign Transformation speaks to another feature of the framework, leveraging sound information to change over talked language into sign dialect representations. This capability improves openness for people with sound-related impedances, empowering them to lock in in conversations and pass on their contemplations utilizing sign dialect motions. Sign to Text Transformation serves as the converse handle, translating hand signals captured through cameras into printed yield. This highlight finds application in circumstances where clients may lean toward to communicate using motions instead of written or talked dialect.

The system's capabilities expand past phonetic communication, including Object detection. This segment utilizes cutting-edge calculations to distinguish and classify objects inside pictures, upgrading the system's utility in recognizing and connection with the encompassing environment. The ultimate segment, Language Identification discourages the dialect utilized within the given input, showcasing the system's versatility to different phonetic settings. Consolidating these capabilities into an all-encompassing outline-work is accomplished through fastidious integration of modules. The measured plan guarantees adaptability, versatility, and effectiveness. By leveraging the qualities of Content Recognition, the capabilities of Firebase ML Pack, and the accuracy of Signal Acknowledgment, the framework conveys an upgraded client encounter that consistently bridges a bunch of input strategies with their comparing yield representations.

The Framework Plan encapsulates a modern integration of six specialized areas, each contributing to an energetic and comprehensive user experience. Through the synergy of Content Acknowledgment, Firebase ML Pack, and Motion Acknowledgment, this plan exhibits our commitment to saddling innovation for exact and flexible dialect and question acknowledgment. By addressing different modes of interaction, we point to form a comprehensive stage that engages clients to communicate viably and exceptionless, notwithstanding of their favoured communication mode or phonetic foundation.

7) IMPLEMENTATION

1. Text to Sign Conversion

The text recognition system created in Android Studio blends classic text input with revolutionary hand gesture recognition, providing users with an interesting and dynamic experience. Users enter text using familiar techniques, which are subsequently translated into equivalent hand motions. Image recognition techniques and machine learning algorithms enable precise gesture recognition. The system converts motions back to words, allowing users to evaluate and change their input. Benefits include dynamic engagement,

individualized gestures, and demonstrating the combination of classic and new interaction approaches. Overall, the system is a unique combination of traditional input with cutting-edge technology that improves user experience and emphasizes the possibility for intuitive interactions with digital devices.

2. Picture to Sign Conversion

The built text recognition system links the Google Vision API with Android Studio, allowing for text extraction from photos as well as novel hand gesture identification. The procedure begins with the extraction of text from photographs using OCR technology. Each character is then assigned a corresponding hand motion, improving user interaction. The Android software takes hand motions in real time, identifies them using image recognition techniques, and then maps them back to characters using machine learning algorithms. This integrated system provides users with convenience, efficiency, and engagement while demonstrating the combination of modern technology and creative interfaces. Overall, it marks a big step forward in user interface and technological integration, simplifying text extraction and improving the user experience through gesture-based input methods.

3. Voice to Sign Conversion

The created text recognition system uses Google's Voice API and Android Studio to transcribe spoken words into text and incorporates hand motion recognition for interaction. The procedure begins by using the Voice API to translate spoken words into text, which acts as the foundation for further interactions. Each character is paired with a hand motion, which improves user involvement. The Android software takes spoken words in real time, recognizes them with speech recognition technology, and maps them to motions using machine learning techniques. This integrated system provides users with convenience, efficiency, and engagement while demonstrating the combination of speech and gesture recognition capabilities and innovative interfaces. Overall, it marks a huge step forward in human interaction and technology convergence, simplifying spoken word conversion and increasing user engagement through gesture- based input modalities.

4. Sign to Text Conversion

The project's goal is to create a Sign-to-Text Conversion system in Android Studio that combines real-time hand motions taken by the device's camera with American Sign Language letters. The system interprets movements using picture recognition and machine learning techniques, which are similar to computer vision principles. The system learns to detect individual letters and converts gestures into text representations on the screen after being trained with an ASL-gesture dataset. This method improves accessibility and communication for sign language users, fostering more inclusive interactions. Furthermore, it demonstrates the convergence of contemporary picture recognition with Android app development, which fosters intuitive communication methods and influences how people interact with technology. Overall, the project emphasizes the relationship between picture recognition, app development, and accessibility, allowing users to efficiently communicate using ASL motions and written text.

5. Object Detection

The project aims to create an Object Detection system using Android Studio, allowing users to record or pick photographs and reliably identify things within them using Google Firebase ML Kit. The technology provides flexibility by allowing users to select between live camera feeds and gallery photographs. The integration of Firebase ML Kit provides robust object detection using advanced image recognition algorithms trained on varied datasets. Detected objects are expressed as a percentage, showing the system's level of confidence in their recognition. This percentage- based representation provides useful insights into image content, improves user experience, and has practical uses in a variety of settings. The project demonstrates the seamless integration of powerful algorithms with user-friendly interfaces, highlighting technology's ability to ease complex processes.

6. Language Identification The project focuses on Language Detection with Google Firebase ML Kit, allowing users to enter text in any language and reliably identify it with the press of a button. The application's text input interface is easy and flexible to a wide range of linguistic preferences. The integration of Firebase ML Kit allows for the examination of entered text to detect its language, using powerful machine learning algorithms trained on a sample of 144 languages. This allows real-time language detection, which improves accessibility for language learners, travellers, and others dealing with multilingual information.

8) RESULTS AND DISCUSSIONS

The results of our Sign language recognition models are shown in Table 1. Our approach involves training a CNN model for converting sign language to text. For text-to-sign conversion, we utilize the HashMap class to map characters to their respective signs. Voice-to-sign conversion relies on the Google Voice API. In the case

of image-to-sign conversion, we initially extract text from the image using the Google Cloud Vision API, followed by converting the text to the corresponding sign.

Feature	Implementation	Accuracy
Sign to Text	CNN (Convolutional Neural Network)	95%
Voice to Sign	Google Voice API	91.34%
Image to Sign	Google Cloud Vision API	92.45%

Table 1: Feature Specific Accuracy

9) CONCLUSION

The study covers potential enhancements to hand gesture recognition systems, including generalizing the system to include more gestures and actions, as well as training the system on data from several users to account for variances in gesture execution. User testing is useful for identifying errors in recognition accuracy, it also discusses the key techniques, applications, and challenges of hand gesture recognition, including gesture acquisition methods, feature extraction, classification, and applications in sign language and robotics. Environmental issues and dataset availability are addressed, emphasizing the need for additional research in the topic. While current methods have demonstrated great performance, there is still opportunity for exploration and growth of hand gesture detection into other technical domains such as tablets, smartphones, and game consoles. Hand gesture recognition has the potential to improve human-computer interactions by making them more natural and pleasurable. The study also introduces an automatic hand-sign language translator for mute/deaf people and discusses system requirements and performance objectives. It goes into detail into software issues such as system startup and recognition algorithms, as well as challenges in identifying ambiguous measurements and recommending technical solutions.

10) REFERENCES

[1] Research paper," A Survey on Hand Gesture Recognition for Indian Sign Language", IRJET, 2016.

[2] Research paper," Gesture Recognition and Machine Learning Applied to Sign Language Translation", Springer, 2016. [3] Research Paper" Towards Interpreting Robotic System for Fingerspelling Recognition in Real Time", ACM, 2018.

[4] Research Paper "Hand Gesture Movement Recognition System Using Convolution Neural Network Algorithm", International Research Journal of Computer Science (IRJCS) Issue 04, Volume 6 (April 2019) ISSN: 2393-9842.

[5] " Hand Gesture Recognition for Sign Language: A New Hybrid Approach", Research gate, January 2020

[6] "Development of an End-to-End Deep Learning Framework for Sign Language Recognition, Translation, and Video Generation", IEEE Access, 2022

[7] "SignBERT: A BERT-Based Deep Learning Framework for Continuous Sign Language Recognition", IEEE Access, Volume 9, Journal Article, 2021

[8] "A Review of the Hand Gesture Recognition System: Current Progress and Future Directions". IEEE Access, Volume 9, Journal Article, 2021

[9] "Deep Learning-Based Approach for Sign Language Gesture Recognition with Efficient Hand Gesture Representation", IEEE Access, Volume 8,Journal Article, 2020

[10]" Real-Time Static Hand Gesture Recognition for American Sign Language (ASL) in Complex Background", SCRIP, 2012