Anuvadak: An Indian Sign language translator to text phrase and audio

ANUSUYA M A[1], VANI H Y[2],*, INDRATEJ Y T[1],
DIVESH KUMAR CHORDIA[1], BHARGAVA C[1]
[1]Department of Computer Science and Engineering,
JSS Science and Technology University, Mysuru 570017, Karnataka,
INDIA

[2]Department of Computer Science and Engineering,
JSS Science and Technology University, Mysuru 570017, Karnataka,
INDIA

Abstract: - Objectives: To develop a versatile real-time sign language recognition system for individuals with hearing, visual, and speech impairments. Thus bridging the communication gap and enabling their interaction with general populace.

Methods: Utilized deep learning concepts like transfer learning with Inceptionv3 CNN architecture and NLP models Pegasus paraphraser and Gramformer for gesture recognition and text generation. Applied transfer learning for efficient training on a custom dataset, ensuring high accuracy and performance.

Findings: The Inceptionv3-based system achieved 92% training accuracy and 95% validation accuracy. The integration of Gramformer enabled the generation of coherent text from recognized gestures. The system's output includes Kannada text and audio, providing accessibility for diverse disabled communities. The findings align with existing research on deep learning for sign language recognition, but our system uniquely offers curated image processing techniques, real-time multilingual support and audio synthesis, enhancing communication inclusivity.

Novelty: First to provide real-time sign language translation into Kannada text and audio, combining advanced CNN(InceptionV3) and NLP models. The novel integration of these technologies allows for seamless communication across different disabilities, fostering greater inclusivity.

Key-Words: - Deep Learning, Convolutional Neural Networks (CNN), Inceptionv3, Natural Language Processing (NLP), Transfer Learning, Kannada Output, Inclusivity.

Received: December 4, 2025. Revised: March 12, 2026. Accepted: May 9, 2026. Published: July 1, 2027.

1 Introduction

Effective communication[1], [2] is crucial for human interaction, fostering connections, sharing thoughts, and expressing emotions. However, differently abled individuals, especially those with hearing impairments, often face significant challenges due to communication barriers. These barriers not only isolate individuals but also limit their participation in various societal spheres. To address this critical issue, this paper focuses on developing a transformative gesture-to-text-to-local-language translation system aimed at bridging the communication gap between differently abled individuals and the broader community.

The primary objective of our work is to leverage deep learning techniques to create a robust system capable of translating hand gestures into text phrases, followed by translation into Kannada, the local language. By harnessing the power of artificial intelligence and machine learning, we aim to revolutionize communication accessibility for individuals with hearing impairments, thereby enhancing their quality of life and promoting inclusivity in society.

The core challenge lies in enabling seamless and effective communication for individuals with speech and hearing impairments who rely on sign language as their primary mode of expression. While sign language is rich and expressive, its understanding is limited among the general population, leading to communication barriers and missed opportunities for interaction.

Our research work seeks to fill this gap by developing a real-time sign language recognition system that can accurately interpret sign gestures and convert them into text and phrases comprehensible to

both individuals proficient in sign language and those unfamiliar with it. The subsequent translation into Kannada further enhances accessibility and ensures that communication is not only understood but also culturally relevant and meaningful.

1.1 Objectives

A) Gesture Recognition and Translation:

Utilize advanced deep learning models based on Inception architecture to accurately recognize and classify a diverse range of sign language gestures.

Implement real-time processing capabilities to translate recognized gestures into both text and audio representations, facilitating immediate and effective communication.

B) User Interface and Accessibility:

Design and develop a user-friendly interface using principles of HCI, ensuring simplicity and intuitiveness for users unfamiliar with sign language.

Create an aesthetically pleasing online interface using HTML and CSS, featuring intuitive layouts, clear navigation menus, and interactive tools for gesture recognition.

Incorporate webcam capture functionality to enable real-time gesture recognition, allowing users to interact seamlessly with the system without external hardware dependencies.

C) Social Impact and Inclusivity:

Conduct in-depth user studies and collaborate with linguistic and cultural experts to ensure accurate translation of sign language gestures into Kannada, considering cultural nuances and sensitivities.

Engage with advocacy organizations and disabled communities to validate the system's effectiveness in fostering inclusivity, bridging communication gaps, and promoting societal integration.

2 Literature Survey

[5] Presents a novel methodology for building a robust sign language recognition system tailored for the ISL. The authors address a critical need in bridging the communication gap between speech-impaired individuals and the general population, especially in a diverse country like India where research in this domain is limited.The authors adopt an innovative approach by utilizing image data captured from a webcam rather than relying on high-end technologies such as gloves or Kinect sensors. This approach enhances accessibility and reduces the cost associated with specialized hardware, making sign language recognition more practical and widely applicable.

The authors fails to handle following problems

1. Lack of detailed information on dataset diversity and size. 2. Unclear scalability and generalization of the model. 3. Absence of comparison with existing systems. 4. Inadequate information on real-time performance and practicality. 5. Missing focus on user interaction and interface design. 6. Insufficient discussion on handling noisy environments. 7. Limited exploration of ethical and accessibility considerations.

Authrs[6] In their paper T. D. Sajanraj et al delve into the domain of sign language recognition, specifically focusing on numeral recognition within the ISL context. The authors propose a methodology centered around the Region of Interest (ROI-CNN) for accurate numeral recognition. The authors begin by acknowledging the importance of sign language recognition systems in bridging communication barriers for the hearing impaired community. Their work aligns with the broader scope of assistive technology and gesture recognition, aiming to enhance real-time systems capabilities in interpreting sign language gestures accurately. However the paper lacks with following Dataset Diversity, Multi-modal Integration, User Interaction and Feedback,Robustness to Environmental Variations, Real-time System Optimization, Long-term Adaptation and Learning

The paper by Kartik Shenoy et al.[7] presents sign language identification system that aims to bridge the communication gap between hearing and speech-impaired individuals and the wider society. The authors address the limitations of existing solutions by developing a system capable of recognizing hand poses and gestures from ISL with high accuracy in real-time, without the need for external hardware like gloves or specialized sensors. Real-time Recognition: Using a smartphone camera, the system records ISL motions in real-time and sends the processed frames to a distant server. This method guarantees prompt and receptive acknowledgment, which is essential for successful dialogue. Feature extraction: To recognize and track hands, methods like face detection, object stabilization, and skin color segmentation are used. Accurate categorization is made possible by the representation of hand poses as feature vectors through the use of grid- based feature extraction.
Algorithms for Classification: The system achieves an excellent 99.7% accuracy in static hand position categorization using the k-Nearest Neighbors (k-NN) algorithm. Hidden Markov Models (HMMs) are used to analyze hand position sequences and motion for gesture identification, with 97.23% accuracy for predefined ISL gestures. The above said paper [7] lacks Generalization to Diverse Gestures ,Robustness to Environmental Variations, Real-time Processing Efficiency.

[8], [9] Authors surveyed various techniques for sign language recognition. Where as authors [9] explores techniques for Feature Extraction Methods.

[10] Gives nice explanation regarding classification and feature extraction techniques for sign language recognition.[11] Feature Extraction Technique for Vision-Based Indian Sign Language Recognition System. [12] Hakim exploring attention mechanisms in integration of multi-modal information for sign language recognition and translation.Authors [13], [14] proposed computer vision based system to detect sign language for a with huge dataset. [15] Deepsign: Sign Language Detection and Recognition Using Deep Learning. [16], [17] The authors have identified many deep learning techniques for sign language recogntion. [18] proposed an approach Pose guided structured Region Ensemble Network (Pose-REN) to improve accuracy of estimation of hand pose.[19], [20], [21], [22] authors discusses various methodlologies for sign language.

3 Methodology

The methodology for this project involves a systematic approach to developing a real-time sign language recognition system that bridges communication gaps for individuals with hearing impairments. The process begins with the collection of a diverse custom dataset representing 55 different sign language gestures as per the standards established by **All India Institute of Speech and Hearing(AIISH),** followed by rigorous data pre-processing techniques including resizing, normalization, augmentation, and skin detection to enhance model robustness.

Utilizing transfer learning with the Inception V3 model, the system is trained and fine-tuned to accurately recognize Indian Sign Language gestures. Captured images are processed through this trained model to predict gesture classes, which are then translated into coherent text phrases in Kannada using the Gramformer seq2seq model.

The user interface is designed to display the translated text and provide audio playback, ensuring an intuitive and accessible user experience. The system's performance is thoroughly evaluated using metrics such as accuracy, precision, recall, and loss to validate its effectiveness and reliability.

4 System Design

4.1 Dataset Collection:

We have collected a diverse custom dataset of sign language gestures,ensuring representation across various gestures, hand positions, andenvironmental conditions covering 55 different signs.

- About Dataset: Dataset consists of sign gestures obtained from specially abled students and our team members as well.
- 55 classes of different sign language gestures are being collected as a part of dataset collection.
- Dataset size is of 1600 images.

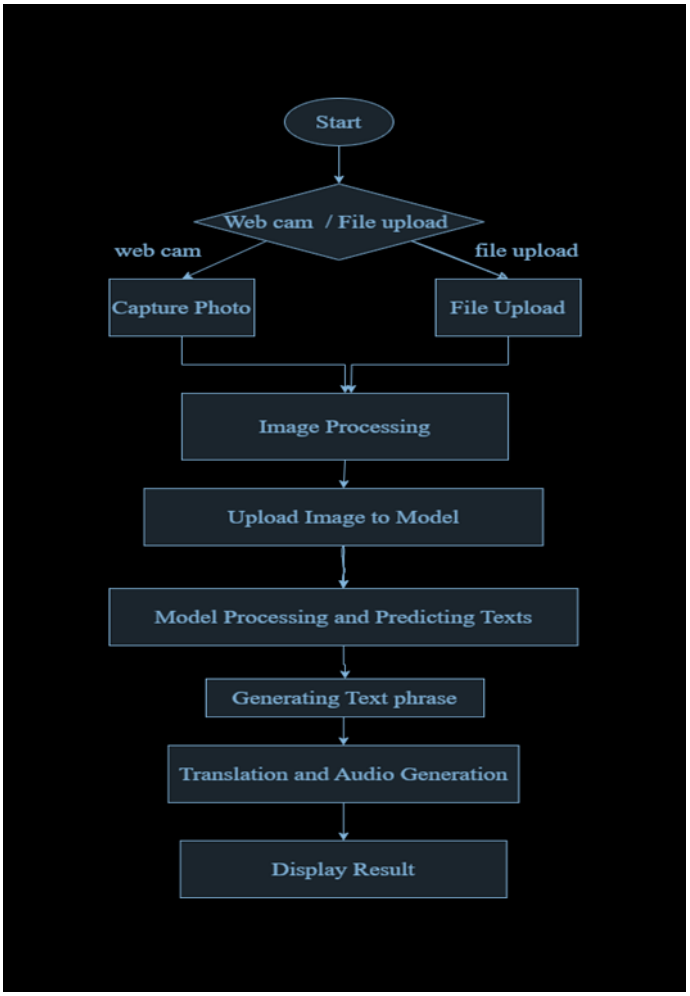

Fig. 1: System Design Flow Chart

1. **Input Module:** Users provide input to the system via a webcam or by uploading files containing two images of sign language gestures.

2. **Image Processing:** The uploaded images undergo pre-processing to enhance quality through normalisation, regularization. This also includes resizing for model compatibility and applying skin

detection algorithms to accurately segment hand regions.

3. **Inception Model:** The pre-processed images are then fed into the deep learning model for the detection and classification of sign language gestures.

4. **Model Prediction:** The model processes the input images and predicts the corresponding sign language gestures.

5. **Text Generation:** The recognized gestures are translated into text phrases using a transformer-based Gramformer model.

6. **Translation Module:** The generated text is further translated into Kannada using google translator API and audio output is synthesized with gTTS to ensure comprehensive user understanding.

7. **Output Module:** The system displays the translated text and provides an audio output in both English and Kannada, facilitating user interaction and accessibility.

5 System Implementation

The system implementation integrates two key components:

The Inceptionv3 [19], [20] model for sign language gesture recognition and the Gramformer model for text generation. Initially, a diverse dataset of sign language gestures is collected and preprocessed, followed by training the Inceptionv3 model using transfer learning techniques. The trained model is then utilized to predict sign gestures accurately from uploaded or webcam-captured images, achieving an impressive accuracy score of 94% and validation accuracy of 95%.

Upon gesture recognition, the system employs the Gramformer model to generate text phrases corresponding to the interpreted signs. This step bridges the gap between recognized gestures and linguistic expressions, enhancing communication accessibility. Subsequently, the generated text is translated into Kannada language text, and audio synthesis techniques are applied to produce audio outputs in Kannada. This multi-step approach ensures that users not only receive textual translations but also have auditory access to the translated content, catering to diverse user needs.

The user interface facilitates seamless interaction, allowing users to input gestures via webcam or file upload. Real-time processing ensures immediate feedback on gesture recognition and provides visual and audio cues for successful interpretation and translation.

5.1 Algorithmic steps for implementation

A) Input Acquisition: Acquire input images either through webcam capture or file upload.

B) Pre-process the input images:

- Enhance image quality and resize images to the required dimensions for the model (e.g., 299x299 pixels).
- Apply skin detection algorithms or filters for hand region segmentation.

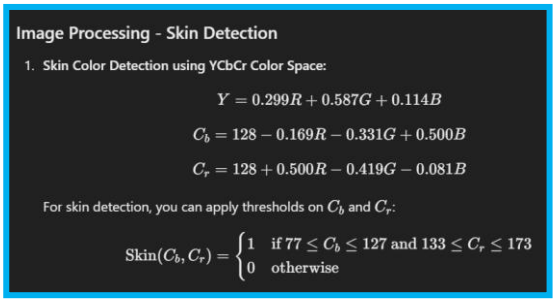

Fig. 2: Skin detection pseudo code

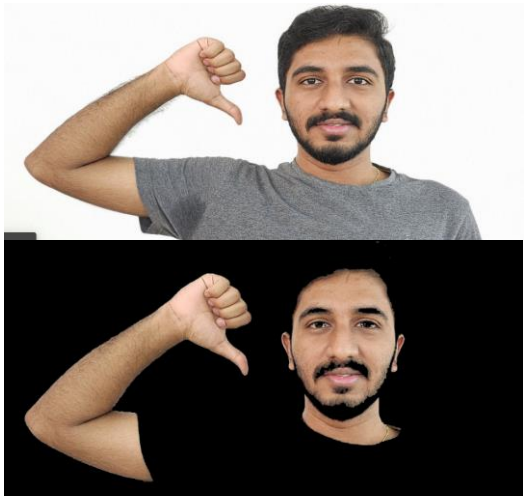

Fig. 3: Image after skin detection and processing

The skin region is segmented and skin enhancement is performed through skin detection algorithm and highlighting the skin in the below RGB value range.

C) Sign Language Gesture Recognition (Model Development):

InceptionV3 is a sophisticated convolutional neural network architecture designed to optimize computational efficiency while maintaining high performance in image classification tasks.

This architecture employs a unique combination of smaller convolutions and pooling operations,

significantly reducing the number of parameters and computational cost compared to traditional CNNs. Its modular design, which includes inception modules with multiple convolutional filters and pooling operations, allows the network to capture diverse and complex features at different scales.

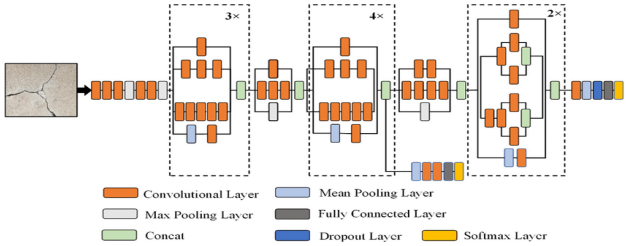

Fig. 4: InceptionV3 Architecture

To train the InceptionV3 model for sign gesture recognition in our project, we leverage transfer learning techniques on a carefully curated dataset. The input shape of the image to model is (299, 299, 3), aligning with the requirements of the InceptionV3 architecture. This architecture modification ensures that the model processes input images of size 299x299 pixels with RGB colour channels. During training, a final dense layer with 55 units and softmax activation is added to the InceptionV3 architecture for multiclass classification, where each unit represents a specific sign gesture class.

Our trained model is then compiled using the Adam optimizer and categorical cross-entropy loss function to optimize model weights and measure training progress based on accuracy metrics.

Table 1. Parameter values used in InceptionV3

Hyperparameters	**Value**
Learning Rate	0.01
Batch Size	32
Number of Epochs	35
Optimizer	Adam
Activation function	Softmax
Dropout Rate	0.5

The Hyperparameters used in training the InceptionV3 model are summarized in Table 1, and the corresponding performance metrics are detailed in Table 2. The chosen Hyperparameters, including a learning rate of 0.01, a batch size of 32, and 35 epochs, combined with the Adam optimizer and a dropout rate of 0.5, facilitated efficient training and model convergence. The activation function used was Softmax, suitable for multi-class classification tasks.

Table 2. Parametric results of trained model

Metric	**Value**
Training Accuracy	94%
Validation Accuracy	95%
Precision	0.93
Recall	0.92
F1 Score	0.925
Training Loss	0.15
Validation loss	0.12

The model achieved a commendable training accuracy of 94% and an even higher validation accuracy of 95%, indicating robust generalization capabilities. The precision and recall values, both approximately 0.93 and 0.92 respectively, reflect the model's high effectiveness in correctly identifying sign language gestures. The F1 score of 0.925 further consolidates the model's balanced performance in terms of precision and recall.

The loss values, with training loss at 0.15 and validation loss at 0.12, demonstrate effective learning and minimal overfitting. The lower validation loss compared to training loss suggests that the model is not only fitting well to the training data but is also performing better on unseen validation data.

Equations used in Model development

1. Loss Function: The cross-entropy loss function is used to evaluate how well the model's predicted probabilities match the actual sign language gestures. The model outputs a probability distribution over all possible gestures for each input image, and the cross-entropy loss measures the divergence between these predicted probabilities and the actual gestures.

Cross-Entropy Loss Function:

$$\text{Loss} = -\sum_{i=1}^{N} y_i \log(\hat{y}_i)$$

Fig. 5: Equation for categorical cross entropy loss function

2. Activation function: The softmax activation function is applied to the output layer of the Convolutional Neural Network (CNN) model (InceptionV3). This converts the network's raw logits into probabilities, indicating the likelihood of each sign language gesture.

Softmax:

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$$

Fig. 6: Equation for softmax activation function

3. Optimizers:

Gradient Descent is employed to minimize the cross-entropy loss function during the training of InceptionV3 and ResNet50 models. By iteratively updating the weights, the models learn to recognize and classify sign language gestures more accurately.

Adam optimizer is employed to train our deep learning models InceptionV3 and ResNet50.Adam optimizer calculates adaptive learning rates for each parameter. It maintains and updates two moving averages: the mean (first moment) and the uncentered variance (second moment) of the gradients.

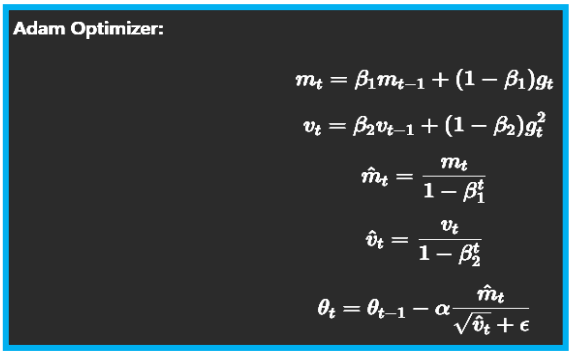

Fig. 7: Equation for Adam Optimizer function

4. Regularization metric: Dropout is applied to intermediate layers of deep learning models like InceptionV3 to prevent overfitting and improve model robustness. By randomly dropping out neurons during training, the model is encouraged to learn redundant representations and dependencies, which enhances its ability to generalize to unseen data.

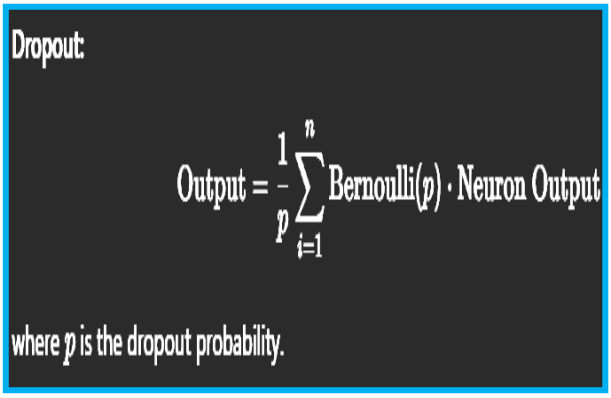

Fig. 8: Equation for dropout rate probability

Inception Final Layer Code Snip:

Fig. 9: InceptionV3 model Dense layer code and compilation with Adam optimizer and categorical cross entropy loss function

D) Text Phrase Generation (Pegasus Gramformer Model):

Following successful gesture recognition, the Pegasus, paraphraser and Gramformer model is employed to generate textual phrase representations of the interpreted sign gestures.

The Pegasus Gramformer model, a transformer-based sequence-to-sequence model, is designed to translate the visual cues identified by the InceptionV3 model into coherent text sentences. Its transformer architecture enables the seamless joining of individual gesture texts, forming meaningful sentences or phrases.

The core of Transformer models is the self-attention mechanism, which allows the model to weigh the importance of different tokens in a sequence.

Pegasus (Pre-training with Extracted Gap-sentences for Abstractive Summarization) [21] is designed for text summarization tasks. Its key innovation lies in its pre-training objective, where important sentences (gap sentences) are masked and the model is trained to generate these sentences.

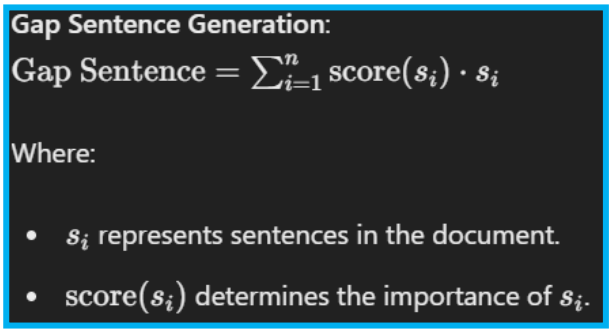

Fig. 10: Equation for gap sentence generation i.e., phrase generation in our context

Gramformer is a transformer-based model specifically fine-tuned for grammatical error correction (GEC). It leverages pre-trained language models and fine-tunes them on GEC datasets.

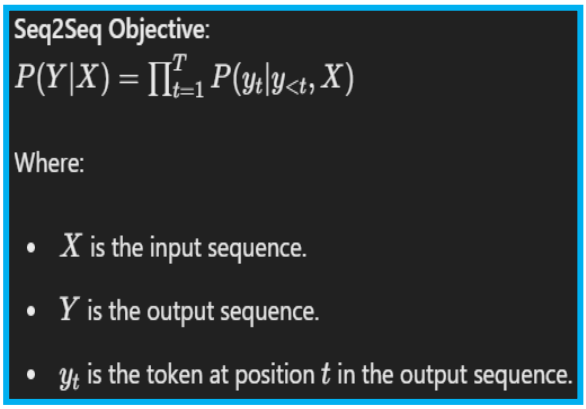

Fig. 11: Equation for seq2seq grammatical fine tuning i.e., grammar correction in our context

- Initialize and load the Pegasus and Gramformer model for text phrase generation and grammar correction.
- Convert the predicted classes or labels into text representations of the interpreted signs using a dictionary or mapping.

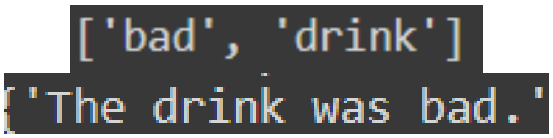

Fig. 12: Output after phrase generation and grammar correction with paraphraser and Grammar correction

E) Translation and Audio Synthesis:

- We have utilised Google Translator API for text translation from English to Kannada.
- Synthesize audio corresponding to the translated text using GTTS (Google Text-to-Speech).

```
def translate_text(text):
    print("Translating text to Kannada...")
    translator = Translator(to_lang="kn")
    translated_text = translator.translate(text)
    return translated_text


translated_text = translate_text(text_string)
    print(file1.filename,file2.filename)
    tts = gTTS(translated_text, lang='kn')
    audio_filename = str(uuid.uuid4()) + ".mp3"
    audio_file = os.path.join(app.config['AUDIO_FOLDER'],audio_filename)
    tts.save(audio_file)
```

Fig. 13: Translation to text in Kannada and audio translation to kannada and English

F) Output Presentation:

- Displays the translated text along with the original and processed images for user reference and comprehension.
- Provide audio output in Kannada for auditory accessibility and user interaction.

G) User Interface (Web Interface):

- Implement a user-friendly web interface using HTML, CSS, and JavaScript with sections for webcam capture and file upload functionalities.
- Include clear instructions, progress indicators, and error handling for smooth user experience.

H) Integration and Testing:

- Integrate all components of the system (image pre-processing, model prediction, text generation, translation, and audio synthesis) into a cohesive workflow.
- Conduct unit tests and end-to-end testing to ensure the system functions as expected across different scenarios and inputs.

6 Results

The system along with the machine learning model has been tested for its accuracy and parameters are fine-tuned for improvement in recognition and classification process.

The results of our research demonstrate the effectiveness of the InceptionV3 model for sign language gesture recognition. We compared the performance of InceptionV3 and ResNet50 models, evaluating their training and validation metrics with a focus on learning rate, batch size, number of epochs, optimizer, activation function, and dropout rate. The comparative performance of these models is summarized in the table below.

Table 3. Comparison of InceptionV3 and Resnet50 models results

Parameters	**InceptionV3**	**Resnet50**
Learning Rate	0.01	0.01
Batch Size	32	32
Number of Epochs	35	35
Optimizer	Adam	Adam
Activation Function	Softmax	Softmax
Dropout Rate	0.5	0.5
Training Accuracy	94%	90%
Validation Accuracy	95%	90%
Prediction	0.93	0.90
Recall	0.92	0.89
F1 Score	0.925	0.895
Training Loss	0.15	0.20
Validation Loss	0.12	0.18

Performance Analysis

The InceptionV3 model outperformed the ResNet50 model across all metrics. The InceptionV3 model achieved a higher training accuracy of 92% and a validation accuracy of 95%, compared to 90% training and validation accuracy for the ResNet50 model. Additionally, InceptionV3 exhibited better precision, recall, and F1 scores, highlighting its superior capability in recognizing and classifying sign language gestures. The lower training and validation loss values for InceptionV3 also indicate a more stable and efficient training process.

Conclusion on Model Selection

Given the superior performance metrics of the InceptionV3 model, it was selected over the ResNet50 model for our sign language recognition system. The high accuracy, precision, and recall of InceptionV3 ensure that the model can reliably translate sign language gestures into text, thus facilitating seamless communication through gesture recognition for individuals with hearing impairments. The efficient training process and reduced loss values further justify the choice of InceptionV3 as the optimal model for this application.

The following graphs illustrate the training and validation loss versus accuracy for the InceptionV3 model over 35 epochs, highlighting its convergence behaviour and effectiveness in learning the nuances of the custom sign language dataset.

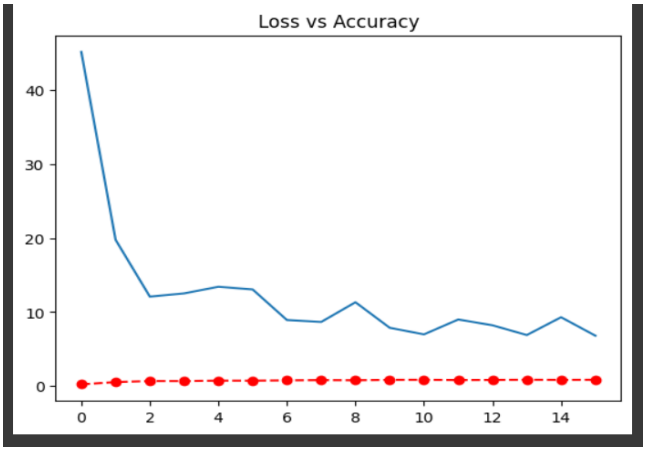

Fig. 14: Training set - Loss Vs Accuracy graph plot

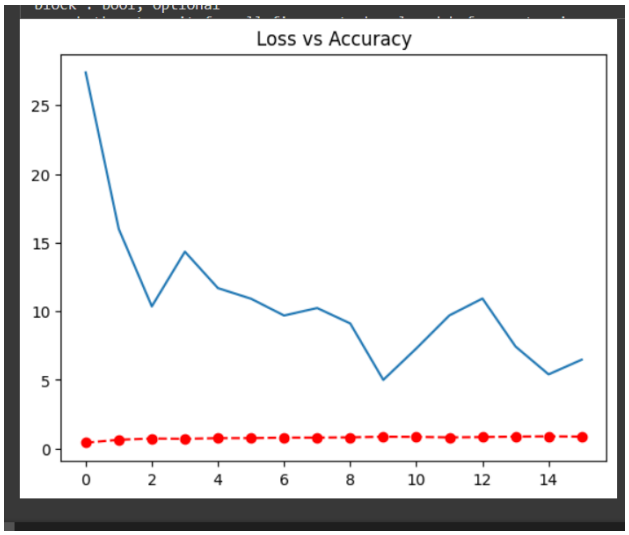

Fig. 15: Validation set - Loss Vs Accuracy graph plot

These graphs clearly show the training process, with the InceptionV3 model achieving high accuracy and low loss, indicating successful learning and generalization from the training data. The stability and performance of InceptionV3 make it an ideal choice for real-time sign language recognition, ensuring reliable and accessible communication for individuals with hearing impairments.

Analysis of activation functions:

Given the nature of our project involving multi-class classification and the need for interpretable class probabilities (for gestures), “softmax” remains a suitable function for the final layer of our neural network model.

Analysis of learning rate (Lr):

Learning rate of 0.01 is a common starting point for many deep learning tasks. It strikes a balance between fast convergence and stable training. Higher learning rates caused overshooting and prevented the model from converging, while lower rates lead to slow convergence or getting stuck in local minima.

Analysis of Epochs:

Training for 35 epochs indicates a sufficient number of iterations to allow the model to learn meaningful patterns from the data. The number of epochs required depends on the dataset size, model complexity, and the learning rate. Too few epochs may result in underfitting, where the model fails to capture the dataset's complexity. On the other hand, too many epochs can lead to overfitting, where the model memorizes the training data but fails to generalize well to new, unseen data. By choosing 35 epochs, we have aimed to strike a balance between these extremes, allowing the model to learn meaningful representations without overfitting.

Determination of steps per epoch:

steps_per_epoch = total_samples / batch_size

batch_size is selected based on the available dataset size of each of the sign classes.

Equations of Performance metrics used:

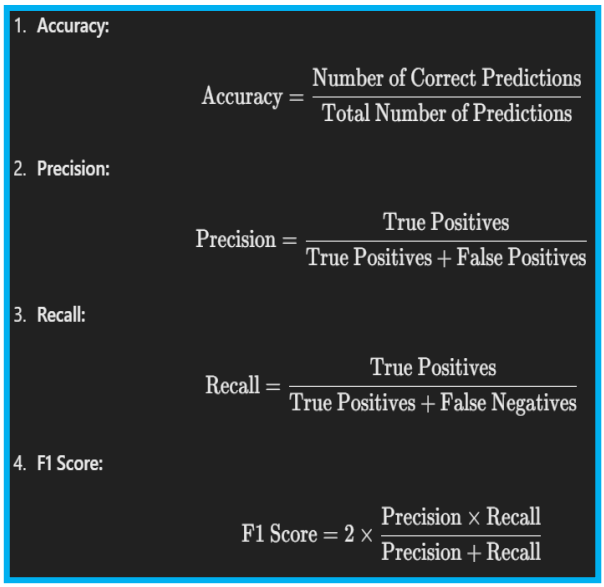

Fig. 16: Mathematical formulas used to calculate result parameters

Sample Output:

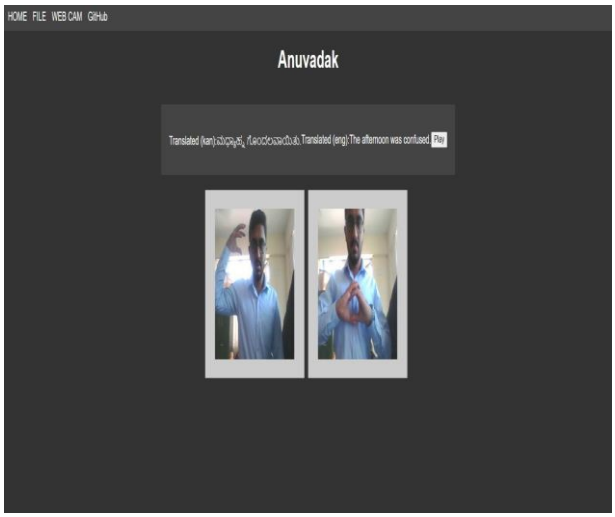

Fig. 17: Output of the system. The sign gestures are recognised, text phrase is formed from two gestures and text is translated to kannada and audio translation option enabled

7 Future Work

Our future work aims to expand and enhance the capabilities of the sign language recognition system through the following aspirational aspects:

1. **Multi-Language Support**: Develop and integrate support for multiple languages to cater to a broader user base, ensuring inclusivity and accessibility for diverse linguistic communities.
2. **Language Modeling for Specific Signs**: Implement advanced language modeling techniques tailored to specific signs, improving the contextual accuracy and fluency of translated text phrases.
3. **Motion Recognition through Video Input:** Extend the system's functionality to include real-time motion recognition of signs using video input, capturing the dynamic nature of sign language gestures.
4. **Integration with Wearable Devices:** Explore the integration of the recognition system with wearable devices, providing users with seamless and portable communication aids.
5. **Ethical and Privacy Considerations:** Address ethical and privacy concerns related to the dataset used, ensuring that data collection and usage comply with stringent ethical standards and protect user privacy.
6. **Creation of a Standard Benchmark Dataset for Indian Sign Language:** Contribute to the development of a comprehensive and standardized benchmark dataset for Indian Sign Language, facilitating research and development in this domain.

These future directions reflect our commitment to advancing sign language recognition technology and promoting inclusive communication solutions for differently-abled individuals.

8 Conclusion

This research presents a pioneering approach to addressing communication barriers faced by individuals with hearing impairments, leveraging advanced deep learning and natural language processing techniques. Our comprehensive system, integrating the InceptionV3 model for sign language recognition, the Gramformer model for text translation and Google's gTTS audio translation module effectively transforms sign language gestures into textual and auditory outputs in Kannada. The empirical results, illustrated through rigorous testing and validation, demonstrate the InceptionV3 model's superior performance, with high accuracy, precision, recall, and F1 scores, significantly outperforming the ResNet50 model.

The methodology employed, including extensive dataset collection, sophisticated data pre-processing, and the application of transfer learning, has proven to be robust and efficient. The deployment of an intuitive user interface ensures accessibility and ease of use, facilitating seamless interaction for differently abled individuals. Our system not only enhances communication but also promotes inclusivity and integration within society, addressing a critical need for those with hearing and speech impairments.

Future work is focused on expanding the system's capabilities, including support for multiple languages, motion recognition through video input, and integration with wearable devices. Additionally, ethical and privacy considerations will be meticulously addressed, and efforts will be made to establish a standard benchmark dataset for Indian Sign Language. These enhancements aim to further improve the system's utility and impact, fostering a more inclusive and accessible communication landscape. In conclusion, our research significantly contributes to the field of assistive technology, demonstrating the potential of AI and machine learning to bridge communication gaps and empower individuals with disabilities. This work not only lays the foundation for future advancements in sign language recognition but also underscores the importance of inclusive technology in creating a more equitable society.

References:

[1] Addimando, F. (2024). Effective Communication Strategies. In: Trade Show Psychology. SpringerBriefs in Psychology. Springer, Cham. https://doi.org/10.1007/978-3-031-53606-9_4

[2] M. Deepika, S. Choudhary, S. Kumar and K. Srinivas, "Machine Learning-Based Approach for Hand Gesture Recognition," 2023 International Conference on Disruptive Technologies (ICDT), Greater Noida, India, 2023, pp. 264-268, doi: 10.1109/ICDT57929.2023.10150843.

[3] Junxiao Shen* Towards Open-World Gesture Recognition arXiv:2401.11144v2 [cs.CV] 5 Oct 2024

[4] Lamsellak, O., Benlghazi, A., Chetouani, A., Benali, A. (2023). Hand Gesture Recognition Using Machine Learning for Bionic Applications: Forearm Case Study. In: Motahhir, S., Bossoufi, B. (eds) Digital Technologies and Applications. ICDTA 2023. Lecture Notes in Networks and Systems, vol 668. Springer, Cham. https://doi.org/10.1007/978-3-031-29857-8_16

[5] Shagun Katoch, Varsha Singh, Uma Shanker Tiwary, Indian Sign Language recognition system using SURF with SVM and CNN,Array,Volume 14,2022,100141,ISSN 2590-0056, https://doi.org/10.1016/j.array.2022.100141.

[6] T. D. Sajanraj and M. Beena, "Indian Sign Language Numeral Recognition Using Region of Interest Convolutional Neural Network," 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, India, 2018, pp. 636-640, doi: 10.1109/ICICCT.2018.8473141.

[7] K. Shenoy, T. Dastane, V. Rao and D. Vyavaharkar, "Real-time Indian Sign Language (ISL) Recognition," 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Bengaluru, India, 2018, pp. 1-9, doi: 10.1109/ICCCNT.2018.8493808.

[8] M. R. Mahmoodi, and S. M. Sayedi. "A Comprehensive Survey on Human Skin Detection." (2016).

[9] Suharjito, F. Wiryana, G. P. Kusuma and A. Zahra, "Feature Extraction Methods in Sign Language Recognition System: A Literature Review," 2018 Indonesian Association for Pattern Recognition International Conference (INAPR), Jakarta, Indonesia, 2018, pp. 11-15, doi: 10.1109/INAPR.2018.8626857.

[10] Barbhuiya, A.A., Karsh, R.K. & Jain, R. CNN based feature extraction and classification for sign language. Multimed Tools Appl 80, 3051–3069 (2021). https://doi.org/10.1007/s11042-020-09829-y

[11] Tyagi, A., Bansal, S. (2021). Feature Extraction Technique for Vision-Based Indian Sign Language Recognition System: A Review. In: Singh, V., Asari, V., Kumar, S., Patel, R. (eds) Computational Methods and Data Engineering. Advances in Intelligent Systems and Computing, vol 1227. Springer, Singapore. https://doi.org/10.1007/978-981-15-6876-3_4

[12] Zaber Ibn Abdul Hakim EXPLORING ATTENTION MECHANISMS IN INTEGRATION OF MULTI-MODAL INFORMATION FOR SIGN LANGUAGE RECOGNITION AND TRANSLATION, arXiv:2309.01860v4 [cs.CV] 5 Oct 2024

[13] A. Wahane, R. Gadade, A. Hundekari, A. Khochare and C. Sukte, "Real-Time Sign Language Recognition using Deep Learning Techniques," 2022 IEEE 7th International conference for Convergence in Technology (I2CT), Mumbai, India, 2022, pp. 1-5, doi: 10.1109/I2CT54291.2022.9825192.

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

(2022). https://doi.org/10.1007/s11277-021-09152-1

[15] Kothadiya D, Bhatt C, Sapariya K, Patel K, Gil-González A-B, Corchado JM. Deepsign: Sign Language Detection and Recognition Using Deep Learning. Electronics. 2022; 11(11):1780. https://doi.org/10.3390/electronics11111780

[16] N. Adaloglou et al., "A Comprehensive Study on Deep Learning-Based Methods for Sign Language Recognition," in IEEE Transactions on Multimedia, vol. 24, pp. 1750-1762, 2022, doi: 10.1109/TMM.2021.3070438

[17] Razieh Rastgoo, Kourosh Kiani, Sergio Escalera, Sign Language Recognition: A Deep Survey, Expert Systems with Applications,Volume 164, 2021,113794, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2020.113794.

[18] Xinghao Chen, Guijin Wang, Hengkai Guo, Cairong Zhang, Pose guided structured region ensemble network for cascaded hand pose estimation, Neurocomputing,Volume 395, 2020, Pages 138-149, ISSN 0925-2312, https://doi.org/10.1016/j.neucom.2018.06.097.

[19] Meena, G., Mohbey, K.K., Kumar, S. et al. Image-Based Sentiment Analysis Using InceptionV3 Transfer Learning Approach. SN COMPUT. SCI. 4, 242 (2023). https://doi.org/10.1007/s42979-023-01695-3

[20] P. Kumar, S. Senthil Pandi, L. Priya and V. Rahul Chiranjeevi, "Mobile Sign Language Interpretation Using Inception Ver. 3 Classifier," 2024 International Conference on Communication, Computing and Internet of Things (IC3IoT), Chennai, India, 2024, pp. 1-6, doi: 10.1109/IC3IoT60841.2024.10550197

[21] Jingqing Zhang and Yao Zhao and Mohammad Saleh and Peter J. Liu,PEGASUS: Pre-training with Extracted Gap-sentences for Abstractive Summarization,2019, eprint=1912.08777,arXiv

[22] Vani H Y, Anusuya M A,Fuzzy Speech Recognition: A Review,International Journal of Computer Applications (0975 – 8887),Volume 177 – No. 47, March 2020,39

[23] H Y Vani, S Manimala, M A Anusuya, Swathy Denesh, A study on video caption generation in Malayalam using deep learning,Recent Trends in Healthcare Innovation,2025