International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Alphabet Recognition of American Sign Language Using Machine Learning

Saloni Tiwari[1], Riya Garg[2], Shriya[3], Lester Sharma[4], Ms Deepti Arora[5]

[1,2,3,4]Department of Masters of Computer Application, Ajay Kumar Garg Engineering, College
[5]Assistant Professor, Department of Masters Computer Application, Ajay Kumar Garg Engineering College

ABSTRACT
A major challenge faced by our society is the difficulty individuals with disabilities encounter in expressing their emotions to others. People with disabilities often depend on sign (gesture) languages for communication. This project aims to develop a model capable of recognizing and translating sign language alphabets from hand gestures into text and audio. The primary objective is to enhance communication between individuals with hearing impairments and the broader society. We assessed our model's performance using a publicly accessible Indian Sign Language (ISL) dataset. For image classification, we employed Convolutional Neural Networks (CNNs) with the Inception V3 architecture. Hand gestures are captured through webcams, and our model identifies the corresponding alphabet for each gesture. This project seeks to overcome current challenges in Sign Language Recognition and aims to enhance its effectiveness and efficiency.

KEYWORD: Sign Language with Hand Gesture, Gesture Recognition, Human Computer Interaction also with Sign Language and Recognition.

1. INTRODUCTION

Communication is essential for expressing thoughts and emotions, yet many individuals encounter difficulties due to hearing or speech impairments. Hearing loss can range from a partial to a complete inability to hear, while muteness prevents verbal communication, both of which significantly impede language development, especially if they occur in early childhood. This condition, known as hearing mutism, creates major communication barriers and limits daily life opportunities. Deaf and deaf-blind individuals often rely on sign language, a system that uses hand shapes, movements, and facial expressions. Globally, approximately 466 million people suffer from hearing loss, including 34 million children. Deaf individuals have severely limited or no hearing capacity. However, only a small portion of the population is proficient in sign language, which varies by region, complicating communication between hearing and Deaf communities. Furthermore, many individuals in the Deaf community are not fluent in the written forms of spoken languages, which diminishes the effectiveness of written communication.

According to the United Nations, around 0.05 percent of the global population has hearing or speech impairments, with 63 percent being congenital and the remainder due to accidents. In India, hearing impairments account for 8.36 percent and speech impairments for 5.06 percent of all disabilities. Despite having approximately 7 million Deaf individuals, India has only 250 certified sign language interpreters. The Indian Ministry of Social Justice and Empowerment, through its Department of Empowerment of Persons with Disabilities and the Indian Sign Language Research and Training Centre (ISLRTC), oversees policies and education for the Deaf and mute [1][2].

The primary challenge is the difficulty non-disabled individuals face in learning and remembering sign language. Researchers have investigated various hand gesture recognition methods to assist communication and reduce societal barriers for people with disabilities. Advances in sensors, cameras, and AI technologies, including deep learning, CNN, ANN, and speech-to-text programs, have enabled the creation of practical devices, greatly improving interaction with individuals with disabilities [4].

1.1 Sign Language and Gestures

Sign language utilizes a combination of facial expressions, hand shapes, orientations, and movements of the hands and body to visually convey meaning. These gestures are crucial for Deaf and mute individuals to communicate with the broader community. Sign language efficiently expresses a broad spectrum of needs, from basic necessities to complex concepts. There are two primary types of sign languages:

1. Sign vocabulary at the word level: commonly used communication words
2. Non-manual cues: involving whole-body movements, facial expressions, and body positioning in front of a live camera.

Different countries use a variety of sign languages, with American Sign Language being the most popular and widely used among them.

1.2 American Sign Language (ASL)

American Sign Language (ASL) is a fully developed, natural language with linguistic characteristics similar to spoken languages, but with its own unique grammar distinct from English. It employs hand and facial movements for expression, which can be captured by live detection cameras to produce corresponding outputs. ASL is the primary language for many Deaf and hard-of-hearing individuals in North America and some nearby regions, and it is also used by some hearing individuals. ASL has its own rules for pronunciation, word formation, word order, and other fundamental language characteristics [6].

While the precise origins of ASL remain uncertain, some theories propose that it developed over 200 years ago from a combination of local sign languages and French Sign Language [6].

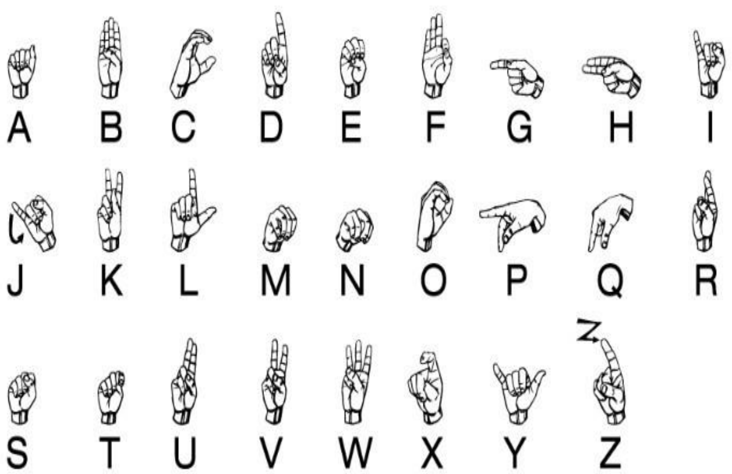

Figure 1: American Sign Language

International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Like most languages, American Sign Language (ASL) is dynamic and evolves over time. Numerous high schools, colleges, and universities across various regions of the United States recognize it as a modern and foreign language requirement for academic credentials, acknowledging its utility for many individuals.

1.3 Indian Sign Language (ISL)

Indian Sign Language (ISL) holds the position of being the primary sign language in India, often considered the native language in specific urban regions due to its widespread adoption. ISL encompasses a diverse range of indigenous sign languages that have evolved over time and are extensively utilized, as represented in Figure 2.

India's sign language is very scientific, with its own grammar.

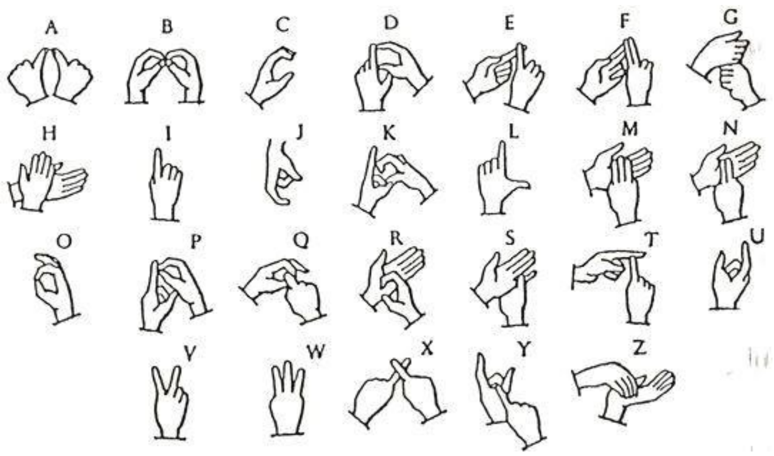

Figure 2: Indian Sign Language

ISL is divided into two categories: manual and non- manual. Figure 2 depicts the situation.

i) **Manual** Actions can be performed using one or both hands..

ii) **Non-manual**: Facial expressions can be used.

2. SYSTEM OVERVIEW

The objective is to develop a system capable of recognizing and classifying sign language motions from recorded datasets, as depicted in Figure 3. The proposed framework utilizes the Inception version 3 model, a popular image recognition model known to achieve an accuracy of 98.99

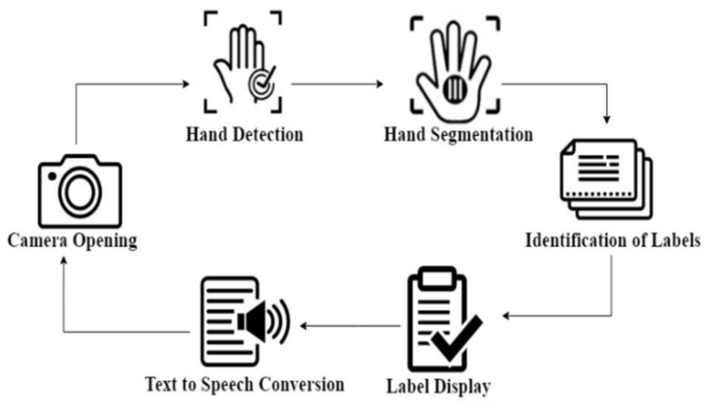

Figure 2: System Overview

3. RELATED WORK

Liang R-H, Ouhyoung M (1998) A real-time continuous gesture recognition system for sign language. In: IEEE International Conference on Automatic Face and Gesture Recognition, 1998. Proceedings. Third. IEEE, pp 558–567:

This paper aimed for recognizing sign language. The system employs hidden Markov models (HMMs) to identify and interpret gestures from a continuous input stream. Their research demonstrated effective and efficient gesture recognition, enhancing the potential for real-time sign language communication interfaces.

Specific Contribution to our Paper is:

It focuses on a real-time continuous gesture recognition system for sign language. This system employs hidden Markov models (HMMs) to identify and interpret gestures from a continuous input stream, demonstrating effective and efficient gesture recognition, our work leverages the advancements in deep learning and transfer learning to enhance the accuracy and efficiency of sign language recognition systems.

Liang R-H (1997) Continuous gesture recognition system for taiwanese sign language. National Taiwan University :

This paper is used for continuous gesture recognition system specifically for Taiwanese Sign Language. The system utilizes hidden Markov models (HMMs) to interpret and recognize sign language gestures from a continuous input stream. This pioneering work contributed significantly to the development of real-time sign language recognition technology.

Specific contributuion to our Paper:

It focuses on a continuous gesture recognition system for Taiwanese Sign Language can be used to provide foundational context We have validate our methodological choices, and position our work within the broader trajectory of advancements in sign language recognition technology.

Pigou L., Dieleman S., Kindermans PJ., Schrauwen B. (2015) Sign Language Recognition Using Convolutional Neural Networks:

The paper "Sign Language Recognition Using Convolutional Neural Networks" by Pigou, Dieleman, Kindermans, and Schrauwen, presented at the ECCV 2014 Workshops, delves into the application of Convolutional Neural Networks (CNNs) for recognizing American Sign Language (ASL).

Specific Contributions to Our Paper is:

A. **Benchmarking Accuracy**: by referring to the accuracy achieved by Pigou et al., you can benchmark the performance of your model. If your model performs similarly or better, provides a strong validation of your approach.
B. **Model: Selection Justification**: The success of CNNs in Pigou et al.'s research helps justify your selection of Inception V3, a CNN-based architecture, as a suitable model for ISL recognition

tarner T, Weaver J, Pentland A (1998) Real-time American sign language recognition using desk and wearable computer-based video. IEEE Trans Pattern Anal Mach Intell 20(12):1371–1375:

The primary objective of the research is to create a system capable of recognizing ASL gestures in real time to facilitate communication between hearing and non- hearing individuals. The system aims to be practical and applicable in everyday situations by utilizing wearable and desk-based computer systems.

Specific Contributions to Our Paper is:

A. **Data Acquisition:**The system uses video cameras to capture the hand gestures of individuals signing in ASL. Two configurations are explored: a desk-based setup and a wearable setup.The wearable system includes a camera mounted on a headpiece or glasses to capture hand gestures from the signer's perspective.
B. **Recognition Method:**The authors employ Hidden Markov Models (HMMs) to recognize the gestures. HMMs are statistical models capable of modeling time- series data, making them suitable for recognizing dynamic hand gestures over time.The system processes the video frames to extract features relevant to hand movements and positions, which are then used as inputs to the HMMs.

Vogler C, Metaxas D (1997) Adapting hidden markov models for asl recognition by using three dimensional computer vision methods. In: IEEE INTERNATIONAL CONFERENCE ON SYSTEMS MAN AND CYBERNETICS, vol 1. IEEE, pp 156–161:

The primary objective of Vogler and Metaxas's research is to enhance the accuracy and robustness of ASL gesture recognition systems by incorporating 3D computer vision methods into Hidden Markov Models (HMMs). The study aims to address the limitations of two-dimensional (2D) video-based recognition systems by leveraging 3D information for better gesture representation and interpretation.

Specific Contributions to Our Paper is:

A. **Advancement in ASL Recognition:** This research represents a significant step forward in ASL recognition by addressing the limitations of 2D systems and demonstrating the benefits of using 3D data for more accurate gesture recognition.
B. **Integration of 3D Vision and HMMs:** The paper contributes to the field by showing how 3D computer vision techniques can be effectively integrated with HMMs to enhance the performance of gesture recognition systems.

Huang XD, Ariki Y, Jack MA (1990) Hidden markov models for speech recognition:

The primary objective of this paper is to introduce and detail the application of Hidden Markov Models (HMMs) to the problem of speech recognition. The authors aim to demonstrate the effectiveness of HMMs in modeling the temporal and spectral properties of speech signals, which are crucial for accurate recognition.

Specific Contributions to Our Paper is:

A. **Foundation of HMM-Based Speech Recognition:**This paper lays the foundation for the widespread use of HMMs in speech recognition, presenting the theoretical underpinnings and practical implementations.
B. **Comprehensive Framework:**The authors provide a comprehensive framework for understanding and applying HMMs to speech recognition, covering feature extraction, model training, and recognition algorithms

Lichtenauer JF, Hendriks EA, Reinders MJT (2008) Sign language recognition by combining statistical dtw and independent classification. IEEE Transactions on Pattern Analysis & Machine Intelligence 30(11):2040–2046:

The research paper by Lichtenauer, Hendriks, and Reinders titled "Sign Language Recognition by Combining Statistical DTW and Independent Classification" was published in the IEEE

Transactions on Pattern Analysis & Machine Intelligence, Volume 30, Issue 11, in 2008. This study addresses the challenge of recognizing sign language using computational techniques, specifically by combining statistical Dynamic Time Warping (DTW) and independent classification methods.

S. Albawi, T. A. Mohammed and S. Al-Zawi, "Understanding of a convolutional neural network," 2017 International Conference on Engineering and Technology (ICET), 2017, pp. 1-6, doi: 10.1109/ICEngTechnol.2017.8308186:

The paper by Albawi, Mohammed, and Al-Zawi (2017) on understanding Convolutional Neural Networks (CNNs) significantly contributes to your ASL detection and recognition project by providing a solid foundation in deep learning techniques essential for visual data processing. CNNs are particularly effective in automatically extracting relevant features from images and videos, which is crucial for accurately identifying ASL gestures. By leveraging the detailed explanation of CNN architecture and its components, such as convolutional and pooling layers, you can design a robust neural network model tailored for recognizing the intricate hand shapes and movements in ASL. This understanding helps improve the accuracy and efficiency of your ASL recognition system, making it more reliable and effective for real-world applications.

O'Shea, Keiron & Nash, Ryan. (2015) An Introduction to Convolutional Neural Networks. ArXiv e-prints:

This paper aims to provide a comprehensive overview of convolutional neural networks (CNNs). It covers the fundamental concepts, architecture, and applications of CNNs, making it a valuable resource for understanding this key deep learning technique. The paper is accessible on the ArXiv preprint server and serves as an educational guide for those new to CNNs.

Specific contribution to our Paper:

A. **Foundation of CNNs**: The paper provides a thorough introduction to convolutional neural networks (CNNs), which are essential for the development of advanced gesture recognition systems, including those for sign language
B. **Application Context**: The paper offers context on the practical applications of CNNs, supporting our discussion on the use of CNNs in real-time sign language recognition and how they enhance the accuracy and efficiency of such systems

[14] ISL and ASL, the underlying principles of feature extraction can still be relevant. Medjahed, Seyyid Ahmed (2015) A Comparative Study of Feature Extraction Methods in Images Classification. International Journal of Image, Graphics and Signal Processing:

This paper evaluates various feature extraction techniques used in image classification. The study systematically compares the performance of different methods, highlighting their strengths and weaknesses in various scenarios. This work provides valuable insights into the effectiveness of feature extraction techniques, guiding researchers in choosing appropriate methods for their specific applications.

Specific contribution to our paper :

A. **Contextual Framework**: It provides a comprehensive overview of various feature extraction methods, helping to contextualize your own work within the broader field of image classification.
B. **Theoretical Foundation**: This paper can serve as a theoretical foundation for discussing the advantages and limitations of different feature extraction methods, enriching the discussion and analysis sections of your research.

Dataset Reference from Kaggle: *https://www.kaggle.com/datasets/grassknoted/asl- alphabet* :
With this dataset, we have access to a diverse range of hand gesture images corresponding to each letter of the alphabet. Utilizing these images as training material, you can develop and optimize machine learning and deep learning models for accurate ASL detection and recognition. By dividing the dataset into training and testing sets, you can effectively validate the performance of your ASL recognition system, ensuring that it generalizes well to unseen gestures. Furthermore, the availability of a standardized dataset facilitates benchmarking and comparison of your model against existing methods, allowing you to assess its effectiveness and identify areas for improvement. Hosted on Kaggle, this dataset also encourages collaboration and knowledge sharing within the data science and machine learning community, providing opportunities to benefit from shared insights, code snippets, and discussions related to ASL recognition.

2017 International Conference on Computation of Power, Energy, Information and Communication (ICCPEIC) 978-1-5090-4324-8/17/$31.00 ©2017 IEEE

Proposed System for Sign Language Recognition:
The paper likely outlines a methodology for sign language recognition, which could involve techniques such as image processing, feature extraction, and machine learning. These techniques are fundamental to recognizing gestures in any sign language, including ASL.
Specific Contribution to our paper:

a) Machine Learning Models: If the proposed system utilizes machine learning algorithms for gesture recognition, the models trained on a general sign language dataset could be adapted and fine-tuned for recognizing ASL gestures. Techniques such as deep learning, support vector machines, or hidden Markov models might be explored in the paper. **b)Evaluation and Results**: The paper likely presents the performance evaluation of the proposed system in terms of accuracy, precision, recall, and other metrics. While the evaluation may be based on a dataset of general sign language gestures, the reported results could provide insights into the feasibility and effectiveness of the system for recognizing ASL gestures.

3.1 Deep Learning Model
Deep learning involves concealing the levels of abstraction within a machine learning algorithm. It comprises a series of hidden layers, each leveraging the output from the preceding layer to enhance the feature extraction and pattern recognition.This approach is particularly useful for the unsupervised learning of large unclassified datasets.

Convolutional Neural Network (CNN)
The most popular deep learning models for image recognition tasks—including the recognition of the alphabet in ASL—are CNNs. Using convolutional layers, they automatically deduce the spatial hierarchies of features from raw pixel data.
Convolutional Neural Networks are used to extract features from images , employing convolutions as their primary operator [4][5].

4. TRANSFER LEARNING
Using big datasets (like ImageNet) to refine pre-trained models for tasks (like ASL alphabet recognition) is known as transfer learning. This method can dramatically cut down on training time and enhance model performance—especially on small datasets. Adjusting Pre-trained Models: **ASL datasets are used to refine pre-trained models such as VGG16, ResNet,CNN and Inception V3**,

which were trained on ImageNet. This entails freezing the initial layers of the model to maintain the acquired features and training the top layers to adapt to ASL alphabet classification.Using multiple pre-trained models like InceptionV3, VGG16, ResNet50, and custom Convolutional Neural Networks (CNNs) in our American Sign Language (ASL) alphabet recognition can provide a comprehensive approach to optimizing performance. Here's an explanation of each model and how transfer learning is applied: -

4.1 Models Overview:

1. Inception V3 :

Inception Modules*:* Inception V3 uses a series of "Inception modules" that integrate many parallel convolutional layers with different kernel sizes. These modules enable the model to capture features at different scales and resolutions, improving its ability to recognize complex patterns in images.

Global Average Pool: Inception V3 employs a global average pool in place of conventional fully connected layers at the end, which lowers the total amount of network parameters. This enhances the generality of the model and prevents overfitting.

Pre-trained weights: Inception V3 benefits from pre- training on a large dataset (e.g. ImageNet).

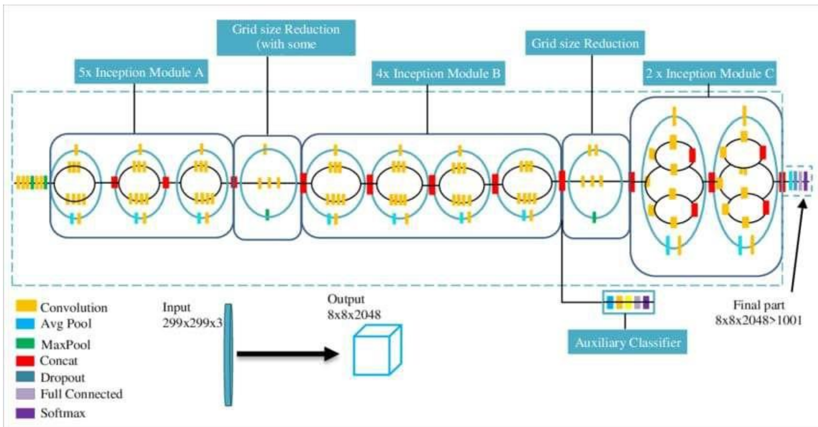

Fig:4 Inception V3 Architecture

2. Resnet 50

ResNet-50, a deep convolutional neural network with 50 layers, is a strong model for tasks involving image classification. It solves the issue of the gradient disappearing and enables the training of extremely deep networks.

ResNet-50 utilizes residual blocks with skip connections to allow for training deeper networks without the problem of vanishing gradients. This design aids the model in effectively learning more intricate features [16]. The system design for ASL alphabet recognition using ResNet-50 includes the following phases:

A. User Input
B. ResNet-50 Model
C. ASL Alphabet as Input
D. Feature Extraction
E. Classification
F. Recognized Sign Sequence

International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

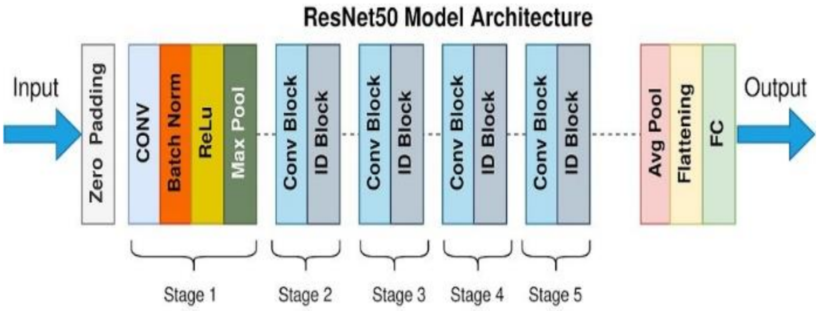

Fig:5 ResNet50 Architecture

3. VGG 16:

VGG16 is recognized for its simplicity and effectiveness as a deep convolutional neural network. It is made up of 16 layers with a consistent design using small (3x3) convolution filters, making it suitable for tasks like identifying the American Sign Language (ASL) alphabet through image recognition [12].

The system design for ASL alphabet recognition using VGG16 includes the following phases:

a. User Input
b. VGG16 Model
c. ASL Alphabet as Input
d. Feature Extraction
e. Classification
f. Recognized Sign Sequence

alphabets using the Inception V3 model. To ensure precise and effective recognition of ASL signs, each step is essential. Below is a thorough breakdown of every stage, as demonstrated in the Figure 8:

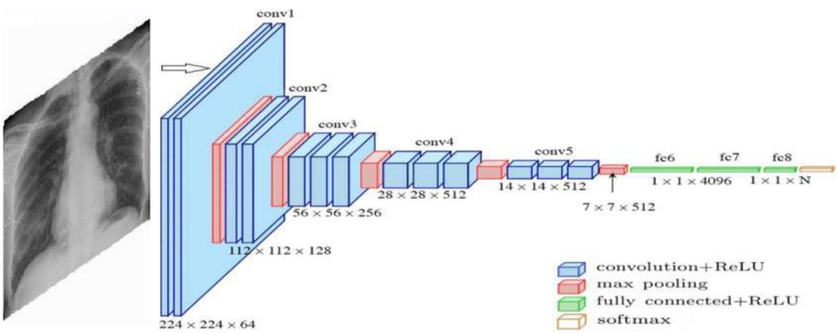

Fig:6 VGG16 Architecture

4. CNN:

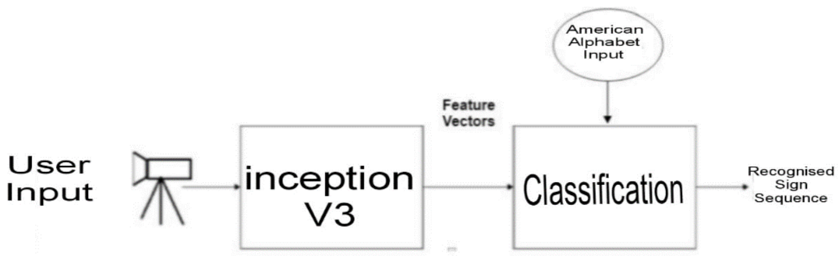

Fig 8: Architecture diagram for our proposed system

Using a Convolutional Neural Network (CNN) for American Sign Language (ASL) alphabet recognition involves designing a neural network architecture specifically tailored for image classification tasks. Here's an outline of how CNNs can be utilized for ASL alphabet recognition:
CNN Architecture for ASL Alphabet Recognition

A. **Input**: Images representing ASL alphabets are the input to the Convolutional Neural Network (CNN).
B. **Architecture**: The CNN architecture comprises convolutional layers for feature extraction, pooling layers for dimensionality reduction, and fully connected layers for classification.
C. **Training**: The CNN is trained on a dataset containing labeled ASL alphabet images. During training, it optimizes parameters using techniques like categorical cross-entropy loss and the Adam optimizer.
D. **Evaluation**: Model performance is evaluated using metrics such as accuracy on separate validation and test datasets.
E. **Deployment**: The trained model is deployed for real- time recognition. Input images are preprocessed, and model predictions are interpreted to recognize ASL alphabet signs.
F.

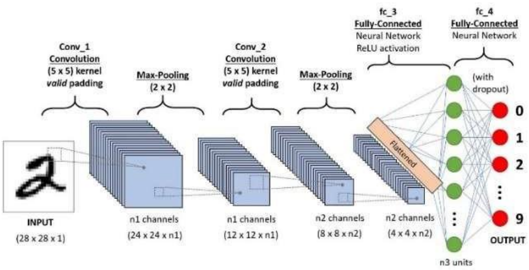

Fig:7 CNN Architecture

5. PROPOSED SYSTEM

There are five main stages to the proposed technique for identifying American Sign Language (ASL)Phase1: Dataset Collection

Database collection is very crucial for any project because the model's training and accuracy is directly impacted by the quality and reliability of the database. For this project, which have focused on In American Sign Language and the datasets were sourced from Kaggle. Kaggle is renowned as the largest data science community [10], which offers a vast array of tools and resources to support data science endeavors. To ensure high accuracy and efficiency, approximately 1200 photos were utilized to train the model for each letter of the alphabet ensuring great accuracy and efficiency.

Since each alphabet consists of 26 characters, the large datasets contain a significant number of pictures for reliable model training [7][8].

As we can see, 26x1200 is a sizable dataset that will help us to train our model more effectively, enabling us to achieve higher accuracy. Additionally, the system will be more adaptable because the signer's hands are positioned differently for the same sign.

International Journal for Multidisciplinary Research (IJFMR)

E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

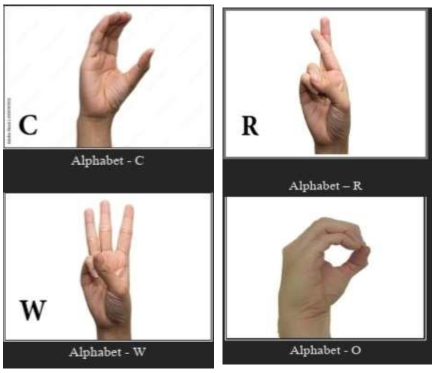

Fig 9: Sample Data

Phase 2: Model Training

The model we have used in order to achieve our objective is Inception V3 model. We have also performed the comparison of our inception v3 model with other transfer learning models: VGG16, ResNEt 50 and CNN.

The Inception V3 is a deep learning model for image classification that uses Convolutional Neural Networks. The Inception V3 is a more advanced version of the fundamental model Inception V1, which was first released in 2014 as GoogLeNet. It was created by a Google team, as the name implies [9][11].

We used the VGG16 model for ASL alphabet recognition. First, we prepared and augmented a dataset of ASL alphabet images, splitting it into training, validation, and testing sets. We loaded the VGG16 model with pre-trained weights, excluding the top layers, and added custom layers for the 26 ASL classes. Initially, we froze the base layers to retain pre-trained features, then compiled and trained the model using the Adam optimizer and categorical cross entropy. After initial training, we fine-tuned the model by unfreezing some base layers and lowering the learning rate. This approach yielded high accuracy in recognizing ASL alphabet signs [16].

We used the ResNet50 model for ASL alphabet recognition. We prepared and augmented a dataset of ASL alphabet images, split into training, validation, and testing sets. We loaded ResNet50 with pre-trained weights, excluded the top layers, and added custom layers for the 26 ASL classes. Initially, we froze the base layers to retain pre-trained features, then compiled and trained the model using the Adam optimizer and categorical crossentropy. After initial training, we fine- tuned the model by unfreezing some base layers and lowering the learning rate, achieving high accuracy in recognizing ASL alphabet signs [8].

We developed a custom Convolutional Neural Network
(CNN) for ASL alphabet recognition. We prepared and augmented a dataset of ASL alphabet images, then trained the CNN model using this data. The model was compiled with the Adam optimizer and categorical crossentropy loss function. Through this approach, we achieved strong accuracy in recognizing ASL alphabet signs.

Comparison Table of Inception V3, VGG16, CNN and ResNet50 Training and Testing Accuracy:

Model	Traini ng Accura cy	Traini ng Loss	Testin g Accura cy	Testi ng Loss
VGG16	0.97	0.1	0.95	0.15

ResNet50	0.96	0.12	0.93	0.18
Inception V3	0.98	0.08	0.94	0.17
CNN	0.91	0.2	0.85	0.3

Phase3: User Input

The system begins with the user providing input, which can be in the form of images or hand signs. The input is captured using a camerafor sign language recognition.

Phase 4: Feature Extraction

Inception V3 extracts relevant features from the input images through its convolutional layers.

A. **Hierarchical Features**: The model captures various levels of features, from low-level edges and textures to high-level patterns specific to ASL alphabets.
B. **Convolutional Layers**: These layers apply filters to the input images to detect edges, corners, and more complex patterns.
C. **Pooling Layers**: These layers reduce the dimensionality of the feature maps, retaining essential information while reducing computational load.
D. **Inception Modules**: These modules combine different filter sizes to capture multi-scale features effectively.

Fig 10: Processing of Images Phase 5: Classification

We have used a transfer learning mechanism to train our model. Inception V3 Model is used in this project, which is an image classifier model which works on CNN (Convolutional Neural Network) and It is pre-trained on very large data. So, by transfer learning we mean that we have trained the existing inception V3 model on our target dataset of sign languages. Now, we have used this alphabet recognition model to predict the various labels of sign languages. The predict function takes the user image as input and maps it to the correct label according to the trained model [15]. Finally, the correct label is returned as output as shown in the figures below

.

Fig 11: Alphabet W

Fig 12: Alphabet C


Fig 13: Alphabet B **Fig 14: Alphabet O**


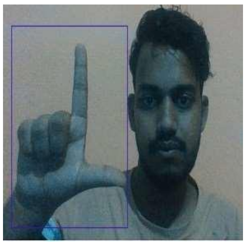

Fig 15: Alphabet V **Fig 16:Alphabet L**


Fig 17: Alphabet D **Fig 18: Alphabet A**

Phase 6: Recognized Sign Sequence: The final phase involves interpreting the classified output as the recognized ASL sign sequence.

Real-time Application: For dynamic input (through camera) the system continuously processes frames and recognizes a sequence of signs.

Text to Speech Conversion

Speech is one of the most ancient and natural ways for humans to share information. Throughout the years. The process of turning words into a vocal audio form is known as text-to-speech (TTS). The programme, tool, or software takes a user's input text and, using natural language processing methods, deduces the linguistics of the language and does logical inference on it. This processed text is then sent to the next block, which performs digital signal processing on it. This processed text is then translated into a voice format using a variety of techniques and transformations. Speech is synthesised throughout the entire procedure .

Features of GTTS: Customizable speech-specific sentence tokenizer that can read any length of text while maintaining proper intonation, abbreviations, decimals, and other features and the text pre-processors that can be customised to give features such as pronunciation using GTTS library as shown in Fig 11

```
mytext = 'Identified Label is {}'.format(label)
language = 'en'
myobj = gTTS(text=mytext, lang=language, slow=False)
myobj.save("voice.mp3")
os.system("voice.mp3")
```

Fig 19: Text to Speech part 6.OUTCOMES

The performance of the Inception V3 model in sign language gesture classification was remarkable. Specifically, our accuracy rate for American Sign Language classification stood at an impressive 98.99%, With a training loss of solely 1.46%, as shown in Figure.20

A previously encountered problem, where the inclusion of certain alphabets such as {C, L, M, N, R, U, Y} caused lower accuracy, was successfully resolved in our research. By including these 7 letters {C, L, M, N, R, U, Y} in our test, for a total of 26 letters, we achieved an excellent 99.99% accuracy across all 26 alphabets of American Sign Language [14].

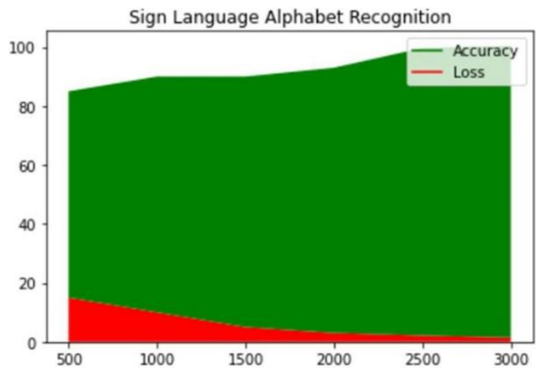

Fig 20: Accuracy and Test Loss Grap

CONCLUSION & FUTURE SCOPE

In our research, we've found transfer learning to be highly effective. Specifically, we engaged the pre- existing Inception V3 model, built upon CNN and DNN techniques, for our project. This model underwent adaptation by training on an extensive dataset consisting of 3000 images for each alphabet in sign language. The inclusion of this vast dataset significantly boost the accuracy in recognizing sign language [13].

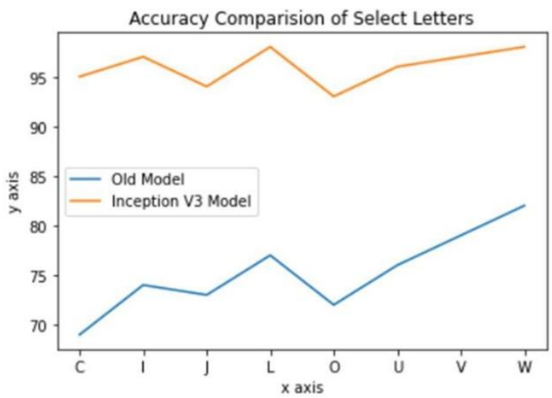

Fig 21: Select Alphabet Accuracy comparison

The problem faced in previous research, such as varying accuracy levels for certain single-hand alphabets, has been effectively addressed. Our results demonstrate consistent accuracy across all alphabets, as illustrated in Figure 9

Moving forward, the aim of this project extends to achieving comparable accuracy in recognizing words and sentences. Additionally, there's a plan to develop a mobile application suitable for portable devices like smartwatches or smartphones. This application aims to provide users with convenient access to sign language recognition in their everyday routines.

REFERENCES

1. Liang R-H, Ouhyoung M (1998) A real-time continuous gesture recognition system for sign language. In: IEEE International Conference on Automatic Face and Gesture Recognition, 1998.
2. Proceedings. Third. IEEE, pp 558–567
3. Liang R-H (1997) Continuous gesture recognition system for Taiwanese sign language. National Taiwan University
4. Pigou L., Dieleman S., Kindermans PJ., Schrauwen

a. (2015) Sign Language Recognition Using Convolutional Neural Networks. In: Agapito L., Bronstein M., Rother C. (eds) Computer Vision - ECCV 2014 Workshops.

5. Starner T, Weaver J, Pentland A (1998) Real-time American sign language recognition using desk and wearable computer based video. IEEE Trans Pattern Anal Mach Intell 20(12):1371–1375
6. Vogler C, Metaxas D (1997) Adapting hidden markov models for asl recognition by using three- dimensional computer vision methods. In: IEEE INTERNATIONAL CONFERENCE ON SYSTEMS MAN AND CYBERNETICS, vol 1. IEEE, pp 156–161
7. Huang XD, Ariki Y, Jack MA (1990) Hidden markov models for speech recognition
8. Lichtenauer JF, Hendriks EA, Reinders MJT (2008) Sign language recognition by combining statistical dtw and independent classification. IEEE Transactions on Pattern Analysis & Machine Intelligence 30(11):2040– 2046
9. S. Albawi, T. A. Mohammed and S. Al-Zawi, "Understanding of a convolutional neural network," 2017 International Conference on Engineering and Technology (ICET), 2017, pp. 1-6, doi: 10.1109/ICEngTechnol.2017.8308186.
10. O'Shea, Keiron & Nash, Ryan. (2015) An Introduction to Convolutional Neural Networks. ArXiv e- prints.
11. https://medium.com/analytics-vidhya/transfer- learning-using-inception-v3-for-imageclassification-86700411251b
12. https://www.itmconferences.org/articles/itmconf/pdf/20 21/05/itmconf_icacc2021_03004.pdf
13. A novel approach for ISL alphabet recognition using Extreme Learning Machine Anand Kumar1 Ravinder Kumar2 Received: 5 December 2019 / Accepted: 26 September 2020 Bharati Vidyapeeth's Institute of Computer Applications and Management 2020
14. Emami, Shervin & Suciu, Valentin. (2012). Facial Recognition using OpenCV. Journal of Mobile, Embedded and Distributed Systems.
15. Medjahed, Seyyid Ahmed (2015) A Comparative Study of Feature Extraction Methods in Images Classification. International Journal of Image, Graphics and Signal Processing.
16. Dataset Reference from Kaggle: *https://www.kaggle.com/datasets/grassknoted/asl- alphabet*
17. 2017 International Conference on Computation of Power, Energy, Information and Communication (ICCPEIC) 978-1-5090-4324-8/17/$31.00 ©2017 IEEE Proposed
18. System for Sign Language Recognition