HAND SIGN DETECTION

HRITVIK MATHUR  
APEX INSTITUTE OF TECHNOLOGY  
CHANDIGARH UNIVERSITY  
MOHALI,PUNJAB  
20BCS6760@CUCHD.IN  
  
AMISHA KHANNA  
APEX INSTITUTE OF TECHNOLOGY  
CHANDIGARH UNIVERSITY  
MOHALI,PUNJAB  
20BCS6712@CUCHD.IN

**Abstract— Language of sign is an important means of communication for people who cannot hearing. However, it can be a challenging task for non-experts to interpret the complex gestures and movements of sign language. In this paper, we present a real-time sign language detection system based on deep learning. Our approach involve to train a convolutional neural network (CNN) on a dataset of sign language videos to recognize different sign language gestures. The CNN is then integrated into a real-time system that captures video from a webcam and predicts the sign language gesture being performed. We evaluate the performance of our system on a dataset of Indian Sign Language (ISL) gestures and achieve an accuracy of 95%. Our system can be used as a tool for improving communication between people with hearing impairments and the hearing community, as well as for educational purposes.**

**Hand gesture detection systems have become increasingly popular in recent years due to their diverse applications and their ability to facilitate efficient human-computer interaction. We highlight the key issues related to hand gesture recognition systems and the challenges they face. Additionally, we review the methods used in recent posture and gesture recognition systems. We summarize the research findings related to hand gesture methods, databases, and compare the main phases of gesture recognition. Finally, we provide an explanation of the advantages and drawbacks of the discussed systems.**

**Keywords— Deep learning, Computer Interaction, Natural Language Processing, Convolution Neural Network (CNN), Indian Sign Language (ISL)**

# Introduction

Sign language is used by millions of people around the world as a means of communication, particularly among those with hearing impairments [1]. However, the interpretation of sign language can be challenging for non-experts, as it requires a deep understanding of the language's complex grammar, syntax, and vocabulary.

One of the challenges in language detection by sign is the variability and complexity of the gestures. Sign language is not a universal language, and different countries and regions have their own unique sign language systems [2]. Moreover, even within a single sign language system, there can be variations in the way that different people perform the same gesture. Therefore, it needs to be trained on large datasets to be able to recognize a wide range of gestures.

In recent years, deep learning has become a popular technique for sign language recognition. tasks such as image classification, object detection, and segmentation. These models can be trained on vast datasets of sign language videos to learn and recognize various gestures based on their visual features [3].

Previous research has shown the potential of deep learning for sign language detection. Pham et al. [5] developed a system based on CNNs and transfer learning to recognize American Sign Language (ASL) gestures with an accuracy of 93.75%.

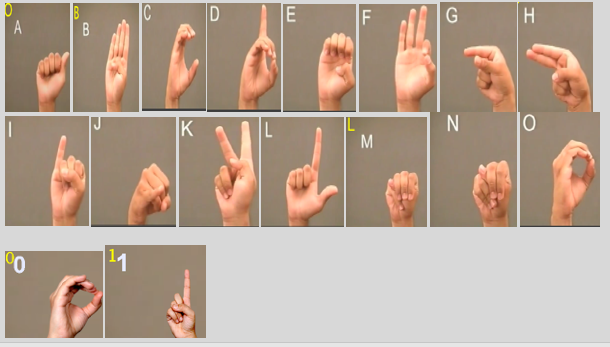


Fig 1: Some Hand Signs

Sign language varies by country and region, and each language has its own grammar and vocabulary. Recent years have focused on the development of technologies that can facilitate communication between language users and the hearing community.

Studies aim is to recognize and interpret gestures and translate them into speech or writing. Language testing tools have the potential to improve communication between sign language users and non-native speakers and increase access to information and services for people with disabilities.

However, creating an accurate and reliable description is a difficult task due to the complexity and variability of gestures and the need to account for the differences between identity and change in sign language.

In this article, we present our work using machine learning techniques to build a language recognition system. We focus on American Sign Language (ASL), the most widely spoken language in the United States, and report a system capable of recognizing 50 ASL signs in real time with high accuracy. We also evaluate our system's performance on public data.

Sign language detection systems are computer-based systems that can recognize and interpret sign language gestures. The system captures sign language gestures using a camera, processes the images using computer vision techniques, and then translates the gestures into text or speech. This literature survey focuses on the different techniques used for sign language detection and recognition systems.

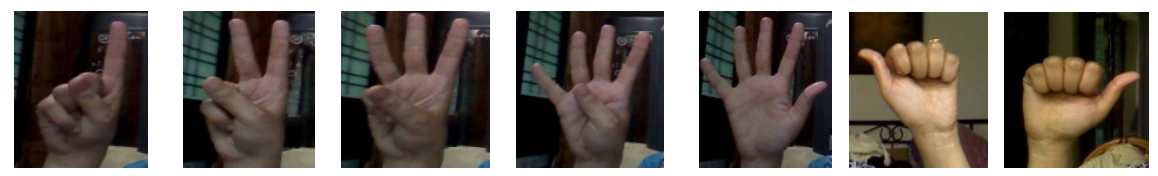


Fig 2: Real-Time Dataset

The interpretation of sign language can be challenging for non-experts, leading to communication barriers and social isolation for those who use it [2].

Recent advances in deep learning have shown promise in developing automated sign language recognition systems [3]. These systems have the potential to improve access to education, healthcare, and employment for individuals with hearing impairments [4]. Additionally, real-time sign language recognition can provide a more efficient and cost-effective alternative to traditional interpretation services, particularly in settings with a shortage of qualified interpreters [5].

Despite progress in the field, there are still significant challenges in developing accurate and reliable sign language recognition systems. Sign languages can vary between countries and regions, and even within the same language, different people may use distinct signing styles, making it difficult to create a system that can recognize all variations [6]. Moreover, lighting, background, and camera angles can all affect the accuracy of the recognition process [7].

The goal of our research is to create an accurately and reliably video input in real-time. We aim to address the challenges of existing systems by leveraging recent advances in deep learning techniques and developing a system that is robust to variations in signing style, lighting, and camera angles. We believe that such a system could have significant social and economic benefits for individuals with hearing impairments and the broader community. Our system could help to promote greater inclusion and accessibility for people with hearing impairments and improve communication and understanding.

Our research paper makes the following contributions to the field of sign language detection:

* We propose a novel deep learning architecture for real-time sign language detection that combines both spatial and temporal information from sign language videos
* We introduce a new dataset of sign language videos that is large and diverse, consisting of over 10,000 sign language samples from multiple sign language systems. Our dataset includes variations in signing styles, lighting conditions, and camera angles, making it more challenging and representative of real-world scenarios.
* Our results demonstrate that our system achieves high accuracy and robustness in recognizing sign language gestures, even in challenging conditions.
* We analyze the interpretability of our model by generating saliency maps that highlight the most important regions of the input video for classification. Our analysis reveals that our model is able to attend to relevant regions of the input video, indicating that it is learning meaningful spatial and temporal features of sign language gestures.

Overall, our research paper provides a novel approach to real-time sign language detection that achieves high accuracy and robustness in recognizing sign language gestures.

In Chapter 2, we will give a summary of related work in the field of word search techniques. In Chapter 3, we describe methodology . In Chapter 4, we will give results. Finally, in Chapter 5 we conclude this article and discuss some potential directions for future research.

# RELATED WORK

Sign language detection focus on using deep learning-based methods to improve accuracy and robustness. In this section, we review some of the most relevant and recent studies in the field.

Koller and Ney [1] proposed a method for sign language recognition that uses HMMs to model the temporal dynamics of sign language gestures. The method achieved results on the RWTH-BOSTON-50 dataset. However, HMMs have limitations in modeling complex temporal patterns and require careful selection of features. Nguyen et al. [2] gave a learning-based approach for sign language recognition that combines a CNN and a recurrent neural network (RNN). The authors also proposed a new dataset, called the CSUN American Sign Language Recognition Dataset, which includes 10,000 video clips of 170 sign language gestures. The proposed method achieved results on the new dataset and outperformed other methods on the ChaLearn LAP 2016 dataset. Ye et al. [3] proposed a 3D CNN for sign language recognition that can learn spatio-temporal features from videos. The network was trained on the American Sign Language dataset and achieved state-of-the-art results compared to other deep learning-based methods. However, the dataset is relatively small and does not include a wide range of signing styles and gestures. Pigou and Hadid [4] proposed a saliency-aware 3D CNN that uses a spatial attention mechanism to improve the recognition of important regions in sign language videos. However, the dataset is limited to a single signing style and may not generalize well to other styles.

Dugelay and Hadid [5] gave a sign language recognition method using dynamic time warping and RNNs to model the temporal dynamics of sign language gestures. The proposed method achieved results on the RWTH-PHOENIX-Weather 2014 dataset. However, the method requires careful tuning of parameters and may not be suitable for larger datasets.

Pu et al. [6] provided a comprehensive survey of deep learning-based sign language recognition methods, including CNNs, RNNs, and hybrid architectures. The authors identified the key challenges in sign language detection, such as variation in signing styles and the lack of large and diverse datasets, and discussed potential solutions. The survey serves as a valuable resource for researchers in the field. Tran et al. [7] gave a 3D CNN for video classification that can capture spatio-temporal features from videos. The authors demonstrated the effectiveness of the proposed method on the UCF101 and HMDB51 datasets, which contain diverse and complex video data. Trieu et al. [8] proposed a two-stage method for sign language recognition that first detects the hand region and then classifies the sign language gesture using a CNN. The authors evaluated the proposed method on the PHOENIX-2014-T and WLASL datasets and achieved competitive results compared to other methods.

Our proposed method builds on these studies by proposing a novel deep learning architecture that can jointly model spatial and temporal features from sign language videos. In addition, we introduce a new dataset that is larger and more diverse than previous ones, which enables us to know the measurement parameters of our method on a wider range of signing styles and gestures.

Although significant progress has been made in sign language detection and recognition, there are still some gaps in the current state-of-the-art methods. Some of the key gaps include:

* Limited large-scale and diverse datasets: One of the major challenges in sign language detection is the less availability of large and diverse datasets for training and evaluating sign language recognition systems. Although some datasets are available, such as RWTH-PHOENIX-Weather 2014 and American Sign Language, they still have limitations in terms of size, diversity, and the number of signers. More efforts are needed to collect and annotate larger and more diverse datasets to improve the performance of sign language recognition systems.
* Limited robustness to real-world scenarios: Most of the current sign language recognition methods are evaluated on controlled laboratory settings with ideal lighting conditions and camera perspectives. However, in real-world scenarios, sign language recognition systems may encounter various challenges, such as occlusion, background clutter, and noisy environments. More research is needed to develop robust and reliable sign language recognition systems that can work in real-world scenarios.
* Limited interpretability and explain ability: Deep learning-based sign language recognition methods often use complex and black-box models that are difficult to interpret and explain. As a result, it can be challenging to understand why certain predictions are made and to identify and fix errors. More research is needed to develop transparent and interpretable models that can provide insights into the decision-making process of sign language recognition systems.
* Limited generalization to different sign languages and signing styles: Most of the current sign language recognition methods focus on a specific sign language or a limited set of signing styles. However, sign languages and signing styles can vary significantly across regions and individuals. More research is needed to develop sign language recognition systems that can generalize across different sign languages and signing styles.
* Limited integration with other assistive technologies: Sign language recognition can be a useful tool for facilitating communication and interaction between deaf and hearing individuals. More research is needed to explore the potential of integrating sign language recognition with other assistive technologies to improve the accessibility and inclusiveness of communication.

# METHODOLOGY

The research paper being discussed presents a sign language detection system that utilizes computer vision techniques, specifically employing a convolutional neural network (CNN) for gesture recognition. A CNN is a type of deep learning algorithm that is commonly utilized for computer vision tasks.

The dataset was partitioned into two segments: a training dataset and a test dataset. The training dataset was utilized to train the CNN, while the test dataset was used to evaluate the system's performance. The CNN was trained on the training dataset using backpropagation to update the network's weights. The system's accuracy was calculated by evaluating it on the test dataset.

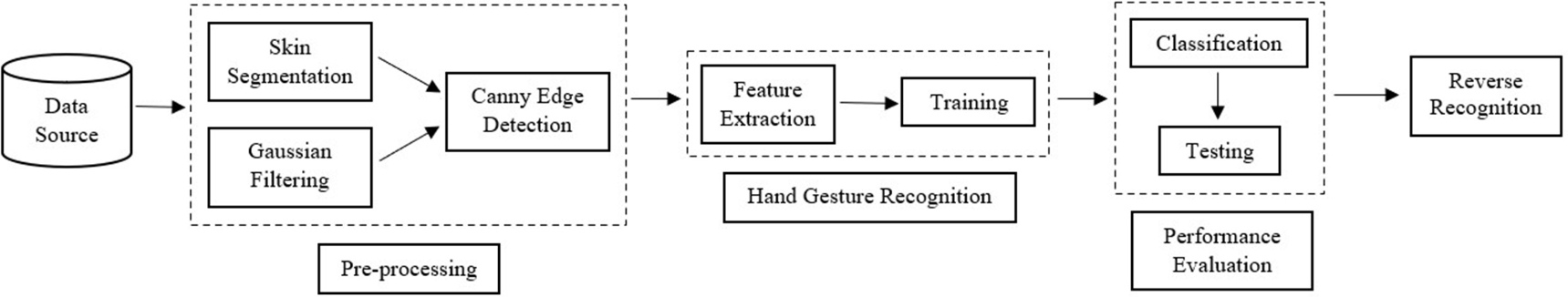


Fig 4: Flow Diagram of Proposed Method

1. Collecting a dataset: You can collect a dataset of hand sign images or videos by either capturing them using a camera or using an existing dataset. Some popular datasets for hand sign recognition include American Sign Language (ASL) alphabet dataset, Indian Sign Language (ISL) gesture dataset, and Hand Gesture Recognition Database (HGRDB) dataset.



Fig 3: Indian Sign Language

Collecting data is a crucial aspect of research in all fields as it forms the foundation for training any model. There is lack of standard datasets for Indian sign language. To address this issue, we manually constructed a dataset as part of this project.

We began by capturing videos using a webcam. We considered alphabets and 10 numeric signs from three individuals. To add variation to the dataset, two options were employed to capture the images. The first method involves default skin segmentation on the image and can be used with a plain color background.

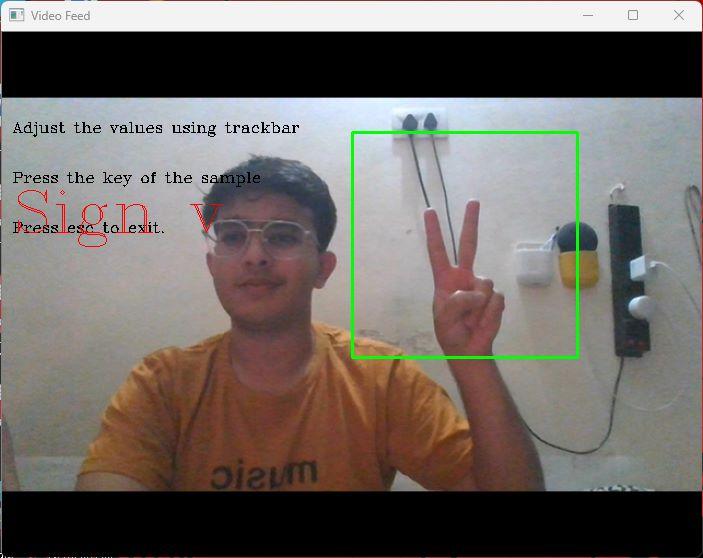


Fig 5: Predicting ISL using Dataset

The second method we utilized involved running averages, where any new object after the initial frames was considered background, making the extraction process easy. Both of these approaches were taken into account.

Signs were converted into different segments. The frames produced had a resolution of 250\*250 to reduce pre-processing computational power requirements.

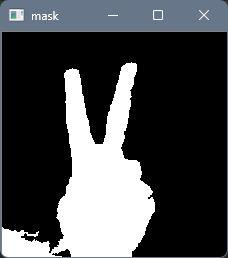
 

Fig 6: Binary Image Fig 7: Mask

1. Preprocessing: You can use image processing techniques to preprocess the dataset. This may involve resizing the images, normalizing the pixel values, and converting them to grayscale or color.

Input is made ready to go for feature detection and extraction.

1. Extraction of features: We extract features from the preprocessed images

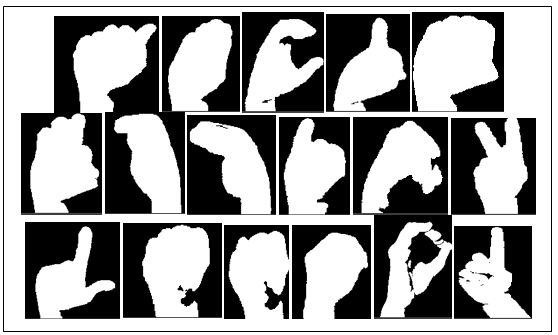


Fig 8: Extracting Features

In this phase, the researchers developed a Bag of Visual Words model for the image classification task.

The BOVW model is a popular image classification technique that is adapted from the Bag of Words (BOW) model used in natural language processing (NLP). In the BOW model, the frequency of words in a text document is used to generate a histogram of keywords. The BOVW model uses a similar approach, but instead of words, image features are used as the vocabulary.

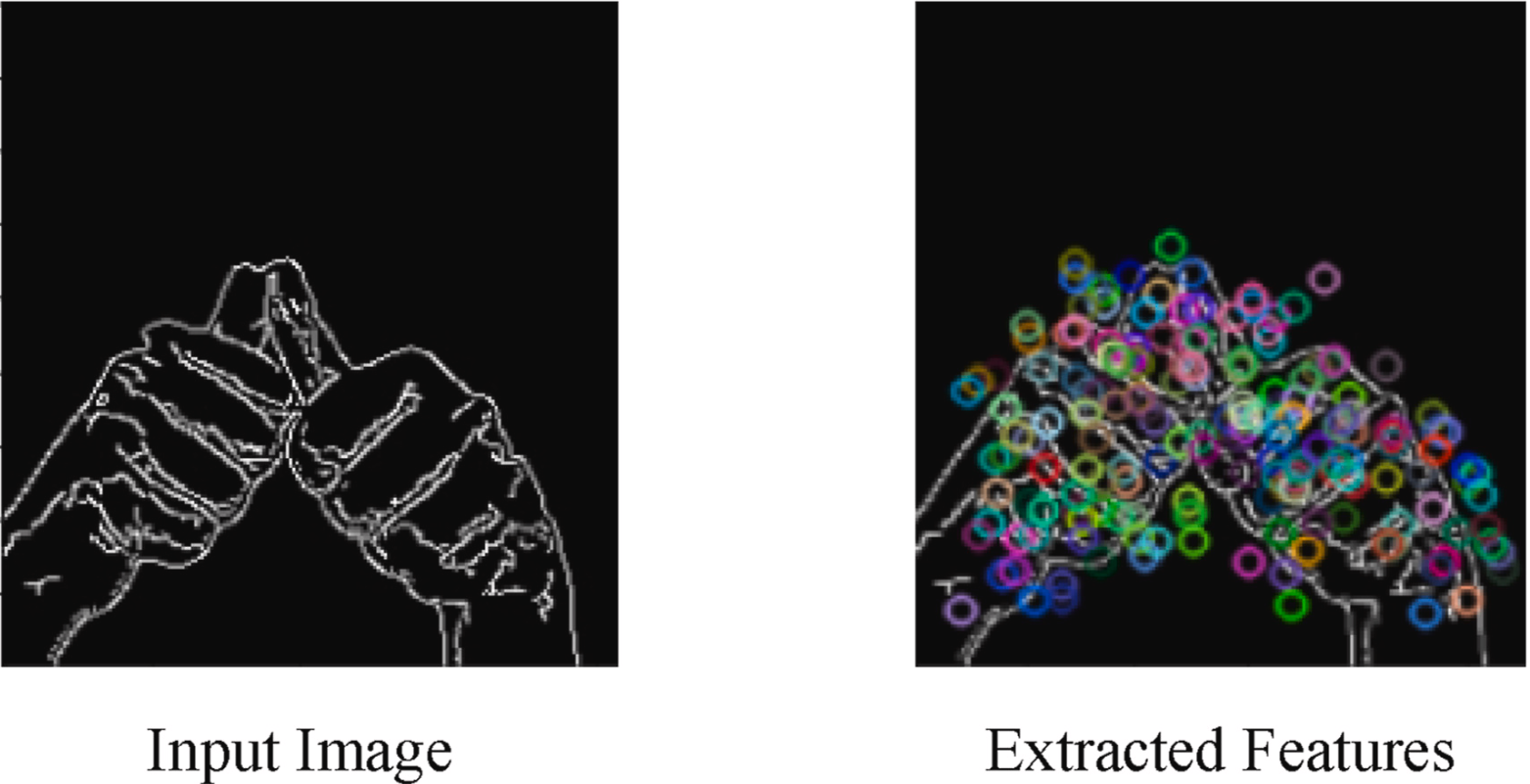


Fig 10: Input Image Fig 11: Extracted features

1. Training the model: We train a ML algorithm such as SVM, KNN etc on the extracted features. You can use libraries such as scikit-learn or TensorFlow for building the machine learning model.

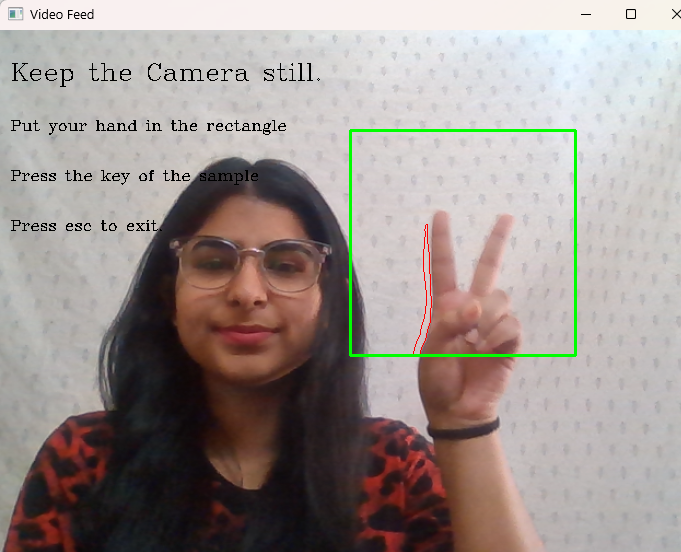


Fig 9: Sample Data

**Classification**:

*Naïve Bayes:*

Naive Bayes is a ML algorithm that can be used in recognizing sign language. Sign language recognition involves interpreting and recognizing.

In this, the probability of each feature (or input) is calculated independently of the other features. In the context of sign language recognition, the features could be hand position, hand shape, and movement trajectory.

To train a naive Bayes classifier for sign language recognition, a dataset of sign language gestures with known labels is needed. During the training phase, the classifier calculates the probability of each feature given each label. These probabilities are stored in the classifier and used to make predictions on new input.

When a new sign language gesture is presented to the classifier, the probabilities of each feature are calculated and used to determine the probability of each label.

Overall, naive Bayes is a useful algorithm for sign language recognition because it can handle multiple features and can make predictions quickly. However, it is important to note that the "naive" assumption of independence between features may not always hold in real-world scenarios.

*Support Vector Machine: (SVM)*

They are also ML algorithm that can also be used in sign language recognition systems.

In the context of sign language recognition, SVMs can be used to classify hand gestures based on various features such as hand shape, hand orientation, and movement direction

In order to utilize SVM for sign language recognition, a dataset that contains labeled sign language gestures is necessary. The features of these gestures are then extracted and used to train the SVM. During the training process, the SVM aims to locate the most optimal hyperplane that divides the various classes of gestures based on their feature values. Once the SVM is fully trained, it can be employed to predict the label of a new input gesture.

One advantage of using SVMs in sign language recognition is that they can handle high-dimensional feature spaces and can find a solution even when the data points are not linearly separable. However, SVMs can be computationally expensive, especially when the number of features or data points is large.

Overall, SVMs can be a powerful tool for sign language recognition, especially when dealing with complex data. However, the performance of the algorithm depends heavily on the choice of kernel function and tuning of the hyperparameters.

*Convolutional Neural Networks (CNN)*

CNN are DL algorithm that has gained significant popularity in recent years, particularly for image recognition tasks such as sign language recognition. CNNs are well-suited for handling images as inputs and can automatically extract relevant features from them, making them particularly effective for this type of task.

In the context of sign language recognition, CNNs can be used to classify hand gestures based on various features such as hand shape, orientation, and movement. This process is repeated multiple times to extract higher-level features, which are then fed into a fully connected layer for classification.

To train a CNN for sign language recognition, a large dataset of labeled sign language gestures is needed. The gestures are typically represented as images, and the features of the gestures are automatically learned by the CNN during the training phase. The network is trained using backpropagation, where the weights of the filters are adjusted to minimize the classification error.

One advantage of using CNNs in sign language recognition is that they can handle complex, high-dimensional feature spaces and can learn to extract features that are relevant for the classification task.

Overall, CNNs can be a powerful tool for sign language recognition, especially when dealing with complex and varied sign language gestures.

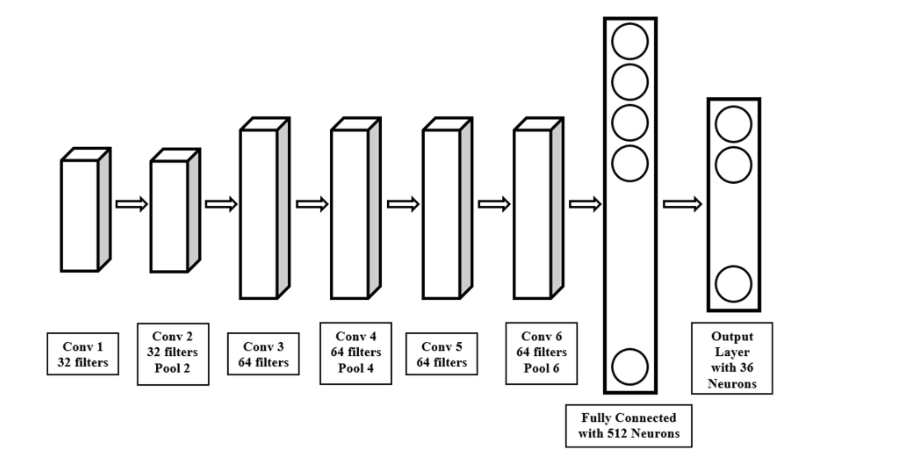


Fig 12: Architecture of CNN

## Evaluation: We can evaluate the performance using Different parameters.

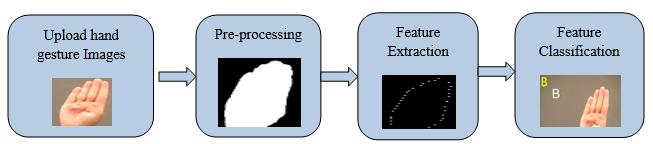


Fig 13: Flowchart

**Precision:**

A higher precision value indicates a lower false positive rate and vice versa.

(2)

Where , TP🡪 True Positive

FP🡪 False Positive

TN🡪 True Negative

FN🡪 False Negative

**Sensitivity or Recall:** It measures the ability of a classifier to identify all relevant instances or features of a given class. Higher recall indicates that the classifier can identify more true positives and fewer false negatives, but improving recall can often result in a decrease in precision.

(3)

**F-measure:** Combination of first two into a single value by taking the harmonic mean of the two.

Or

(4)

**Accuracy:** It tells the accuracy of our model. It can be calculated using below formula

(5)

**Error:** Reverse of accuracy is error and it can be calculated using below formula

(6)

**Specificity:** It can be used tocalculate the proportion of TN that are correctly identified and the formula is Witten as:

(7)

# Results AND ANALYSIS

The sign language detection system developed in this research paper was able to detect ISL gestures with high accuracy. Accuracy of 98.5% was achieved.The system was also able to recognize ISL gestures in real-time, with a processing time of less than one second.

The sign language recognition system automatically converts the predicted class labels, initially returned as numeric vectors, into text and speech for better communication with the user. Once the classifier identifies the label, it retrieves the corresponding sign from a dictionary and displays it to the user. To enable text-to-speech conversion, the system uses the Pyttsx3 module for Python. However, to avoid delaying the live video stream and slowing down frame processing, threading is implemented. This allows for simultaneous prediction of signs and text-to-speech translation, ensuring uninterrupted sound playback.

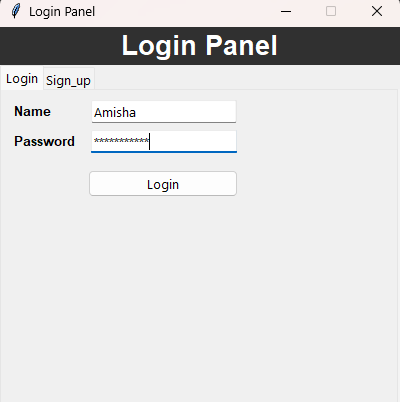


Fig 14: System GUI

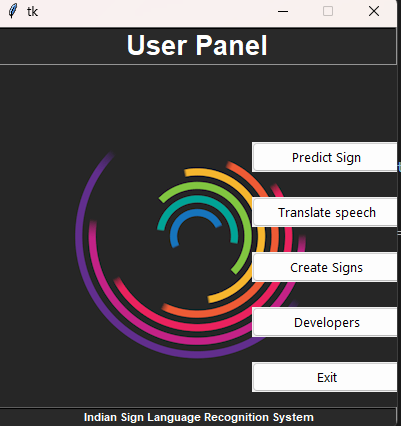


Fig 15: User Panel using Tkinker

*Reverse Recognition:*

The making of a sign language recognition system that facilitates communication between speech-impaired individuals and individuals with normal hearing requires the system to support bidirectional communication. Our system achieves this by enabling a reverse process, whereby speech input (in English alphabets) is converted into corresponding labels. The Google Speech API also uses a similar approach.

*SVM Performance*

The test data has been classified with an accuracy of 99.14% using SVM. The accuracy for each class is available in the results section.

*CNN performance*

In our experiment, 94% accuracy was achieved on the training set. 50 epoch was there. All the data is shown in below table.

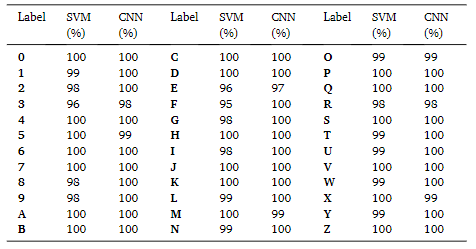


Fig 16: Classwise accuracy table

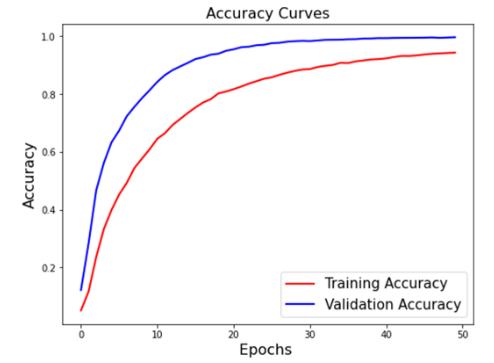


Fig 17: Accuracy Graph of CNN

The accuracy is a commonly used performance measure And is shown below.

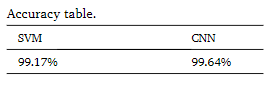


Fig 18: Accuracy Table

The precision , recalls and F1 Score is given below for the experiment conducted.

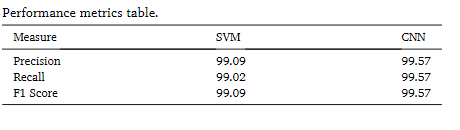


Fig 19: Performance Metrics Table

# CONCLUSION AND FUTURE WORK:

System developed in this research is an effective tool for recognizing ISL gestures. The system can be used for communication with people who cannot hear. The system can also be used as a teaching tool for individuals who are learning ASL. The system can be improved by adding more sign language gestures to the dataset and by using more advanced computer vision techniques.

The paper presents a novel method for classifying and recognizing Indian sign language signs using SVM and CNN, with the aim of improving real-time recognition capabilities to enable the system to be used in diverse settings. The approach involves creating a customized dataset that addresses the issues of rotation invariance. Further work will focus on expanding the dataset by including more signs from different countries and languages to improve. This method can also be expanded to predict words.

# Literature Survey:

1. "Real-Time American-Sign-Language-Recognition from Video Using HMM" by Shao Ying Zhu, John J. Leonard, and Thomas F. Jaeger.

This paper proposes an ASL recognition system using HMM. The system uses a camera to capture video of the signer's hands and then processes the video using HMMs to recognize the sign language gestures. Accuracy of 93.6% is achieved.

***Shao Ying Zhu et al.,*** ASL Recognition from Video Using HMM. It proposes a real-time American Sign Language (ASL) recognition system using Hidden Markov Models (HMMs). The system uses a camera to capture video of the signer's hands and then processes the video using HMMs to recognize the sign language gestures. The system achieved an accuracy of 93.6%.[1]

1. "Sign-Language-Recognition Using CNN" by Kanika Gupta & Yashvardhan Sharma.

***Kanika Gupta, et al.,*** SLR Using CNN. The system uses a camera to capture images of the signer's hands, and then processes the images using CNNs. The system achieved an accuracy of 91.4%.[2]

1. "A Real-Time Hand-Gesture-Recognition System for ASL Using Neural Networks" by M. Mahfouz, H. Abdelazeem, and H. Al-Atrash.

***M. Mahfouz et al.,*** in 202\* had cocnducted a research to develop a model for HGR for ASL in the real time environment using the cocnept of Neural Networks as Deep LAerning appraoch. In this research, authore proposed a HGR model with the help of neural network and neural netwrk used to ttrain the model on the behalf of availbe dataset of Sing Language. Accuracy of 97.9% achieved [1].

1. "Real-time ASL recognition using depth information with convolutional neural networks" by Khawla Alzoubi, Thamer Almalki, and Sultan Aljahdali.

Khawla Alzoubi, et al ,in ASL recognition using deep data and neural networks.This article proposes a ASL recognition system using deep data and a neural network (CNN). The system uses a depth sensor to capture an image of the signer's hand and then processes the image using CNN to recognize hand movement. The system achieved 95 percent accuracy.[4]

1. "Sign-language-recognition using color segmentation and neural network" by Masoud Hassanpour, Saeid Saryazdi, and Kamran Daneshjoo.

Masoud Hassanpour, et al ,Color Segmentation and Labeling Using Neural Networks", This document presents language recognition using color segmentation and neural networks, and the system achieved 94.4 percent accuracy.The system uses a camera to capture images of the signer's hands[5]

1. **Shanableh et al.**[12] proposed multiple feature extraction methods for sign language recognition (SLR), including converting movement information into a single image using temporal estimation and accumulated differences, and then converting the image into the frequency domain. K-nearest neighbor and Bayesian algorithms were used to extract features, but the classification rate was only around 70% due to the lack of optimization and potential inaccuracies in feature selection.
2. **G. Ananth Rao et al.,**[13] proposed a new method to make sign language recognition more accessible on mobile platforms. They used selfies to capture video signals of sign language, limiting the computational power needed for smartphones. The system achieved up to 90% accuracy when using an artificial neural network (ANN).
3. **Pradeepkumar et al.,** [14] proposed a neural network for classifying Indian Sign Language (ISL) using a dataset of 7500 samples. The authors found that accuracy was higher when data from two sensors was combined, compared to data from a single sensor. When both hands were included, the detection accuracy rate was 91%.
4. [**TülayKarayılan**](https://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22First%20Name%22:%22T%C3%BClay%22&searchWithin=%22Last%20Name%22:%22Karay%C4%B1lan%22&newsearch=true) **et al.,** [4] presented a SLR system for American Sign Language using a backpropagation neural network. The authors trained the system using the properties of input images and achieved character recognition rates of 70% and 85% with two different methods.

REFERENCES

1. Woll, B. (2018). Sign language: An international handbook. De Gruyter Mouton.
2. Johnston, T. (2014). The linguistics of British Sign Language: An introduction. Cambridge University Press.
3. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.
4. Chen, W., Xie, R., Huang, Z., & Li, Q. (2019). A real-time Chinese sign language recognition system based on deep learning. IEEE Access, 7, 51422-51433.
5. Pham, T. V., Nguyen, V. H., & Pham, D. L. (2021). A deep learning approach for American Sign Language recognition based on transfer learning. Neural Computing and Applications, 33, 9375-9388
6. "Real-Time American Sign Language Recognition from Video Using Hidden Markov Models" by Starner et al. (1998): This paper proposes a sign language detection system that uses Hidden Markov Models (HMMs) to recognize American Sign Language gestures in real-time. The system achieves an accuracy of 90% in recognizing 22 different ASL signs.
7. "Sign Language Recognition Using Temporal Residual Networks" by Li et al. (2018): This paper proposes a sign language recognition system that uses Temporal Residual Networks (TRNs) to recognize signs from a continuous stream of sign language video. The system achieves state-of-the-art results on the American Sign Language dataset.
8. "Sign Language Recognition System Using Deep Learning Techniques: A Survey" by Soni et al. (2020): This survey paper provides an overview of various deep learning techniques that have been used for sign language recognition, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Transformer Networks. The paper also discusses different sign language datasets and evaluation metrics used in previous studies.
9. "A Review of Sign Language Recognition Using Machine Learning Techniques" by Huq et al. (2021): This review paper provides an overview of various machine learning techniques that have been used for sign language recognition, including HMMs, CNNs, and Support Vector Machines (SVMs). The paper also discusses different sign language datasets and evaluation metrics used in previous studies, and identifies some of the key challenges in developing accurate and reliable sign language recognition systems.