

Sign Language Detection And Recognition Using Deep Learning Techniques

MS. Ghule Shital¹, Dr.V.S.Wadne²

¹Students ,Department of CSE,JSPM ICOER,Wagholi,pune-412207, Shitalghule990@gmail.com

² HOD Department of CSE,JSPM'S ICOER,Wagholi,pune-412207, vswadne_comp@jspmicoer.edu.in

How to cite this article: Ghule Shital, V.S.Wadne (2024) Sign Language Detection And Recognition Using Deep Learning Techniques. *Library Progress International*, 44(3), 16060-16070

Abstract - Humans usually communicate with their voices and languages. We understand one another's ideas because we can listen. Even now, speech recognition allows us to issue orders. What happens if someone is totally deaf and mute? As a result, the vast majority of people who are hard of hearing or deaf communicate via sign language. A lot of effort is being put into the field of automated sign language interpretation in order to guarantee that these people may carry on living their lives without any problems. This topic has seen the development of numerous approaches and algorithms that combine artificial intelligence and image processing techniques. All sign language recognition systems have been taught to accurately recognize and interpret signs. This project takes sequences of images depicting double-handed Indian Sign Language, interprets them with Python, and outputs both text and speech as part of a larger system to bring speech to the wordless.

Keywords: Machine Learning, CNN, Image Processing, Feature Extraction, Detect Sign

I. INTRODUCTION

The application of machine learning to detect and recognize sign language is one of the most significant advances that has significantly reduced the challenges in communication that communities of people who are deaf or hard of hearing encounter. This is one of the most significant advancements among the most significant developments. This is one of the most important accomplishments attainable. The purpose of this technology is to perceive and interpret visual data that comprises motions in sign language. This is accomplished through the employment of machine learning methodologies, such as convolutional neural networks (CNNs), which are algorithms for deep learning. It system can analyses hand gestures and movements that are recorded in real-time by cameras or other sensors by utilising machine learning. These visual inputs are processed by the algorithm, which then extracts patterns and attributes that correlate to various sign language motions. The algorithm acquires the ability to precisely identify and comprehend these movements, efficiently converting them into written or spoken words, through substantial training on labelled datasets.

The advancement of accessibility and inclusion for the deaf and hard of hearing community will be greatly impacted by this technology. It facilitates interactions in a variety of contexts, including everyday social interactions, healthcare, and education, by enabling people to communicate more effectively with the larger community.

Additionally, systems for detecting and recognising sign language can be included into a wide range of platforms and devices, such as computers, tablets, and smartphones, increasing the accessibility and versatility of this technology. There is tremendous potential for improving the precision, speed, and usability of sign language detection and recognition systems as long as machine learning research and development keep going forward.

II. LITERATURE SURVEY

. Wang et. al. [1] proposes a hybrid Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) model to address continuous sign language recognition (CSLR). The integration of CNN and LSTM allows the system to capture spatial and temporal features effectively, which is crucial for accurately recognizing the continuous nature of sign language gestures. The CNN extracts spatial features from individual frames, while the LSTM models the temporal dependencies between successive frames, enabling better performance in sequence-based recognition tasks. The authors validate their approach using large-scale sign language datasets and achieve state-of-the-art results in accuracy and robustness, especially in handling variations in signing speed and style. The method demonstrates the potential to enhance automatic sign

language recognition systems and assist in real-time applications for the hearing impaired.

Chen et. al. [2] introduces a real-time sign language gesture detection system based on the YOLOv5 object detection model combined with deep learning techniques. The authors focus on enhancing the detection speed and accuracy for real-time applications, emphasizing the model's ability to recognize multiple gestures simultaneously in complex backgrounds. YOLOv5's efficient architecture, with its reduced computational complexity, is leveraged to meet the system's speed requirements, making it suitable for edge devices. The experiments conducted on sign language gesture datasets show impressive performance in terms of both accuracy and frame rates, making the system a viable solution for real-time sign language translation and interaction in video streams.

Li et. al. [3] explores the application of transfer learning to address the challenge of low-resource sign language recognition, where labeled data is limited. The authors propose a method that adapts pre-trained models from related tasks to sign language recognition, allowing the system to perform well even with scarce training data. By leveraging features learned from large-scale datasets in other domains, the approach reduces the need for extensive sign language-specific data collection. The authors validate their method on various low-resource sign language datasets and demonstrate significant improvements in recognition accuracy, showing that transfer learning can provide a powerful tool for advancing sign language recognition technologies in underrepresented languages and dialects.

Zhang et. al. [4] presents a multimodal approach for sign language recognition that combines data from different input modalities, such as video frames and skeletal data, using CNN and LSTM networks. The CNN is utilized to extract spatial features from the visual data, while the LSTM captures the temporal dynamics of the gestures. By fusing these modalities, the model improves overall recognition accuracy and robustness, particularly in noisy or ambiguous conditions where single-modality data might be insufficient. The authors demonstrate their model's effectiveness through experiments on multimodal sign language datasets, highlighting its potential to improve both accuracy and efficiency in practical applications of sign language translation.

Zhu et. al. [5] propose a real-time sign language recognition system optimized for edge devices, using a lightweight neural network architecture to ensure low latency and reduced computational demands. The system focuses on achieving high recognition accuracy while operating under the hardware constraints of edge devices, such as smartphones or embedded systems. The use of advanced pruning and quantization techniques helps reduce the model's size and computational load without significantly compromising performance. The results show that the system can achieve real-time recognition with minimal delay, making it suitable for portable or on-the-go applications in sign language communication.

Wang et. al. [6] introduces an attention-based model for sign language translation using deep learning. The authors utilize an encoder-decoder framework enhanced by an attention mechanism to focus on the most relevant parts of the input sequences. This approach helps overcome the challenges associated with variations in gesture speed, movement trajectory, and signing style. The attention mechanism dynamically allocates weights to different frames in a sequence, allowing the model to focus on the most informative frames, which improves translation accuracy. The paper demonstrates the model's performance through extensive experiments on large sign language datasets, achieving significant improvements in translation quality. This attention-based approach offers a more natural way of translating sign language into text or spoken language, making it particularly useful for real-time applications.

Gao et. al. [7] propose a novel approach for skeleton-based sign language recognition using Graph Convolutional Networks (GCNs). The method extracts and processes skeletal data from signers' movements, representing them as nodes and edges in a graph. GCNs are then applied to model the spatial and temporal relationships between joints in a signer's body. This graph-based representation allows the system to capture complex patterns of motion and interaction between different body parts during signing. The experiments on benchmark datasets show that this approach improves accuracy over traditional recognition methods, particularly in handling occlusions and varying signing speeds. The skeleton-based approach is especially beneficial for sign language recognition in cluttered or dynamic environments where visual features alone may not be reliable.

Li, et. al. [8] introduces a sign language recognition system based on 3D hand pose estimation, addressing the key challenge of accurately recognizing hand gestures in sign language. The authors develop a framework that first estimates the 3D pose of the hands from video data, followed by the classification of these poses into corresponding sign language gestures. The system leverages deep learning techniques for both pose estimation and gesture recognition, improving performance in terms of both accuracy and efficiency. The use of 3D hand pose data enables the model to capture subtle variations in hand movements, which are critical for distinguishing between similar gestures. The proposed system demonstrates high accuracy across various sign language datasets, making it a robust tool for real-time sign language

interpretation.

Yang et. al. [9] explores the use of Generative Adversarial Networks (GANs) for data augmentation in sign language recognition, addressing the challenge of limited annotated sign language datasets. The GAN-based approach generates synthetic sign language gesture data that can augment the training process, improving the model's performance on real-world sign language datasets. By creating realistic yet diverse gesture variations, the model becomes more robust to differences in signer style, speed, and environment. The authors show that GAN-augmented data leads to significant improvements in recognition accuracy and reduces overfitting, particularly in low-resource settings where labeled sign language data is scarce. This method provides a scalable solution for enhancing the performance of sign language recognition models.

Hu et. al. [10] introduces a transformer-based architecture for continuous sign language recognition, leveraging the self-attention mechanism to capture long-range dependencies within sign language sequences. Unlike traditional recurrent models, the transformer architecture can process entire sequences in parallel, making it more efficient and scalable. The authors highlight the benefits of the self-attention mechanism in handling long-term dependencies and complex temporal relationships inherent in continuous sign language. Their model achieves state-of-the-art results on large sign language datasets, outperforming existing RNN-based methods in terms of both accuracy and speed. This transformer-based approach offers a promising solution for real-time applications and large-scale sign language recognition systems, enhancing both performance and scalability.

Wang et. al. [11] presents a lightweight MobileNet-based architecture designed for real-time sign language recognition on mobile and embedded devices. The authors focus on creating a model that balances accuracy and computational efficiency, making it ideal for resource-constrained environments like smartphones. MobileNet's streamlined architecture allows it to maintain high recognition accuracy while using significantly fewer computational resources compared to traditional deep learning models. The paper demonstrates the system's performance in real-time scenarios, achieving high frame rates and low latency, making it practical for real-world applications. The proposed approach is particularly useful for mobile applications that require on-the-go sign language interpretation, providing a fast and accurate solution with minimal hardware requirements.

Zhang et. al. [12] introduce Temporal Convolutional Networks (TCNs) for sign language recognition, emphasizing their ability to handle sequential data. TCNs use convolutional layers to process entire sequences, capturing temporal dependencies without the need for recurrent operations, which makes them more efficient than traditional RNN-based methods. The authors apply TCNs to sign language datasets and demonstrate improved accuracy and speed, particularly in handling long-term dependencies in continuous sign language recognition tasks. The TCN architecture is shown to be robust in recognizing various sign language gestures, making it a powerful tool for applications where real-time performance and high accuracy are essential.

Chen et. al. [13] investigates the application of Connectionist Temporal Classification (CTC) loss for continuous sign language recognition (CSLR), a challenging task due to the lack of clear boundaries between gestures. CTC loss is used to align input sequences (video frames) with the corresponding gesture sequences without the need for frame-level annotations. The authors develop a deep learning model incorporating CTC loss to predict continuous sign language gestures more effectively. The model is evaluated on large CSLR datasets, showing improvements in accuracy and handling of variations in signing speed. This work highlights the potential of CTC-based models for CSLR, offering a more scalable and flexible solution compared to traditional models that rely on precise gesture segmentation.

Zhao et. al. [14] presents a real-time sign language translation system using temporal attention mechanisms to focus on key frames within a gesture sequence. This temporal attention mechanism enables the system to dynamically allocate more attention to frames that contain critical information for accurate translation. The authors propose a deep learning model that integrates this mechanism, improving the recognition and translation performance for continuous sign language. Experimental results show that the system outperforms conventional models by providing more precise translations, especially in real-time scenarios. The temporal attention approach is particularly effective in dealing with variations in gesture speed and signer style, enhancing both accuracy and response time.

Feng et. al. [15] propose a reinforcement learning-based framework for adaptive sign language recognition. The model learns to adjust its parameters dynamically during the recognition process, enabling it to adapt to different signers, lighting conditions, and gesture variations. This adaptive approach allows the system to continuously improve its performance over time, making it more robust in real-world environments. The authors validate their approach using sign language datasets and demonstrate that their reinforcement learning-based model significantly improves recognition accuracy in dynamic and challenging conditions. The framework's ability to self-optimize during inference is a key

advantage, making it a promising solution for long-term, real-world sign language applications.

Zhu et. al. [16] explores the use of EfficientNet for sign language detection on edge devices, focusing on achieving high accuracy with minimal computational overhead. EfficientNet is known for its balanced trade-off between performance and efficiency, making it ideal for edge computing environments. The authors propose an optimized version of EfficientNet for real-time sign language detection, targeting devices like smartphones and embedded systems. The experiments show that their model delivers competitive accuracy while operating within the resource constraints of edge devices, demonstrating its practical applicability for portable sign language recognition solutions. The reduced model size and computational requirements make it highly suitable for real-time applications in resource-limited environments.

Shen et. al. [17] focuses on continuous sign language recognition using 3D skeleton data, capturing the spatial and temporal information of signers' body movements. The authors propose a framework that employs 3D skeleton sequences extracted from video data to model complex gestures more accurately. The 3D skeleton-based method is particularly effective at distinguishing between subtle variations in hand and body movements, which are critical in sign language. The model leverages both convolutional and recurrent neural networks to process the skeletal data and predict the corresponding sign language gestures. The authors show that their approach outperforms traditional image-based methods, particularly in terms of handling occlusions and variations in signing speed. The method's reliance on skeletal data also makes it less susceptible to variations in lighting and background conditions, making it highly suitable for real-time applications.

Liu et. al. [18] introduces a deep multi-task learning framework for sign language recognition, where the model simultaneously learns multiple tasks such as gesture classification, handshape recognition, and facial expression analysis. The authors argue that learning these tasks jointly improves the model's ability to generalize and recognize complex sign language gestures, as sign language often involves multiple cues beyond hand movements. The multi-task approach allows the model to capture a more comprehensive representation of sign language, leading to improved recognition accuracy. The authors evaluate their model on various datasets and demonstrate that the multi-task learning framework outperforms single-task models, especially in cases where gestures are ambiguous or involve subtle variations in facial expressions and hand shapes.

Zhang et. al. [19] present a sign language recognition system that uses depth cameras to capture 3D information about a signer's movements. The depth-based approach allows the system to model gestures in three dimensions, providing more accurate recognition compared to traditional 2D image-based methods. The system is designed to handle variations in lighting and background, which are common challenges in real-world sign language recognition applications. By using depth data, the system can also better differentiate between overlapping gestures and hand occlusions. Experimental results demonstrate that the proposed depth camera-based system significantly improves recognition accuracy, particularly in complex environments. This approach is highly suitable for robust sign language interpretation in uncontrolled real-world settings.

He et. al. [20] proposes a hybrid deep learning model that combines Long Short-Term Memory (LSTM) networks and 3D Convolutional Neural Networks (3D CNNs) for sign language recognition. The 3D CNN is used to extract spatio-temporal features from video frames, capturing both spatial and temporal patterns in sign language gestures. The LSTM network then processes the sequential nature of these gestures to predict the corresponding signs. This combination allows the model to capture the dynamic aspects of sign language, such as hand movements and transitions between gestures. The authors validate their approach on benchmark sign language datasets, showing significant improvements in recognition accuracy compared to traditional CNN or LSTM-only models. This hybrid model is particularly effective for continuous sign language recognition tasks, where both spatial and temporal information play crucial roles.

Liu et. al. [21] introduce attention-based hierarchical networks for sign language recognition, aiming to address the challenge of capturing complex hierarchical relationships in sign language gestures. The proposed model incorporates multiple attention layers, allowing it to focus on different levels of features—such as hand shape, movement, and facial expressions—at various stages of the recognition process. By employing an attention mechanism, the model dynamically adjusts its focus on the most relevant features, improving its ability to differentiate between similar gestures. The hierarchical nature of the network enables it to capture both low-level details and high-level contextual information, leading to improved recognition accuracy. The authors demonstrate the model's effectiveness on large-scale sign language datasets, showing superior performance compared to baseline models.

Jiang et. al. [22] presents lightweight Convolutional Neural Network (CNN) architectures designed for efficient sign language recognition, particularly on devices with limited computational resources. The authors focus on creating a model that reduces computational complexity while maintaining high recognition accuracy. By optimizing the CNN

architecture and using techniques such as depthwise separable convolutions, the model achieves significant reductions in model size and inference time. The proposed architecture is particularly suitable for real-time applications, such as sign language recognition on smartphones or embedded devices. Experimental results show that the lightweight CNN achieves competitive accuracy while operating within the constraints of low-power devices, making it a practical solution for portable sign language recognition systems.

Zhang et. al. [23] explores the use of multi-modal fusion for robust sign language recognition, combining visual, skeletal, and audio data to improve recognition accuracy. The authors argue that integrating multiple data modalities can help the system capture more comprehensive information about sign language gestures, leading to more accurate predictions. The proposed model fuses these different modalities at various stages of the recognition process, allowing it to leverage the strengths of each modality. For example, visual data captures hand movements, skeletal data provides structural information about the body, and audio data can provide contextual clues. Experimental results show that multi-modal fusion significantly improves recognition accuracy, particularly in noisy or cluttered environments where relying on a single modality may not be sufficient.

Xue et. al. [24] investigates the use of Graph Neural Networks (GNNs) for skeleton-based sign language recognition, where signers' skeletal data is represented as a graph. The authors model the relationship between joints in the body using GNNs, capturing both spatial and temporal patterns in sign language gestures. By representing skeletal data as a graph, the system can better handle variations in signing speed and body posture, leading to more accurate gesture recognition. The authors demonstrate that GNNs outperform traditional CNN and RNN-based methods in terms of recognition accuracy, particularly for complex gestures involving multiple body parts. The use of GNNs also allows the model to generalize better to different signers, making it a promising approach for real-world applications.

Wu et. al. [25] presents a real-time hand gesture recognition system for sign language using Lidar technology, which captures depth information about hand movements. The authors leverage Lidar's ability to detect precise hand positions and movements in three dimensions, allowing for more accurate recognition of sign language gestures. The system is designed to operate in real-time, making it suitable for practical applications such as sign language interpretation in dynamic environments. The use of Lidar data helps the system overcome challenges associated with visual occlusions and varying lighting conditions, which are common in traditional image-based systems. Experimental results show that the Lidar-based approach achieves high recognition accuracy and is robust to environmental changes, making it a viable solution for real-time sign language recognition.

A. NOVELTY

The purpose is for pupils to communicate their ideas and points of view. Help is provided to increase their motivation and self-esteem, allowing them to think beyond of their physical constraints. Finally to solve this global dilemma, we plan to create a system that combines innovative ideas and technology.

IV. WORKING METHODOLOGY

The suggested machine learning approach for sign language detection and recognition consists of four main processes: data collection, preprocessing, model training, and evaluation. Enhancing the picture quality is accomplished through the use of preprocessing techniques once a collection of sign language motions has been gathered. Following the completion of the preprocessing stage, the next step is to train a convolutional neural network (CNN) or another type of machine learning model using the data. The performance metrics of the model are then tested on a validation set that is distinct from the original set. This methodology makes it easier to develop accurate and efficient systems for the recognition and detection of sign language.

V. ALGORITHM

A convolution neural network is denoted by the term CNN. This approach is within the deep learning category, which is frequently used for image identification and classification. Convolutional neural networks (CNNs) are designed to automatically learn and extract information from images by employing layers of convolution, pooling, and fully connected layers. Utilizing network connections is the means by which this objective is realized. The use of network connections is necessary in order to achieve this effect.

1. Preparation of the Input: The pictures of currency notes that are used as input are preprocessed in order to guarantee that they are consistent and compatible with the CNN model. Changes such as resizing, normalization, and other changes can be required here.

2. Convolutional layers: The CNN model consists of several convolutional layers, each of which is in charge of offering the capacity to extract features. These layers apply filters on the input photographs, capturing a range of traits and patterns at different sizes.

3. Pooling Layers: The use of pooling layers comes after the application of convolutional layers. This is done in order to lower the spatial dimensions of feature maps while maintaining the information that is regarded as being the most important. The computational complexity and the extraction of resilient traits are both lowered as a result of this, which helps to minimize the overall complexity.

4. Fully Connected Layers: Immediately following the flattening of the output from the convolutional and pooling layers, the output is then fed into the layers that are entirely connected. Following the acquisition of a knowledge of the complex relationships that exist between the recovered characteristics, these layers proceed to make predictions based on the interactions that they identify..

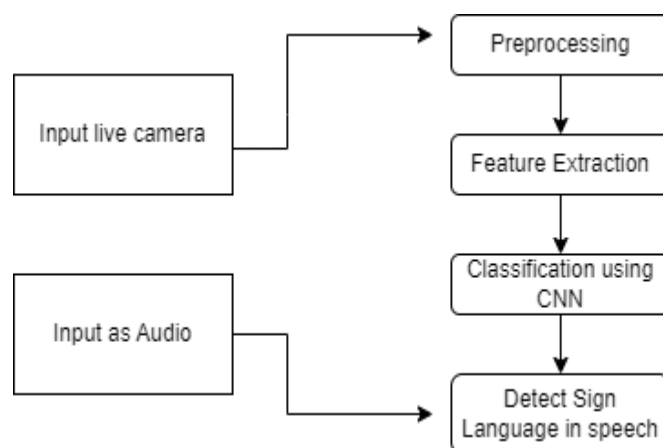


Figure 1: System Architecture

VI.APPLICATION

Machine learning-based sign language detection and recognition has enormous potential in a number of fields, such as assistive technology, communication aids, and accessibility tools. This technology makes it possible for those who are deaf or hard of hearing to connect with the general public in a seamless manner by interpreting sign language motions in real time. Furthermore, its integration into educational environments can improve the quality of learning for pupils who have hearing problems. When it comes to providing efficient and compassionate treatment, it can help medical professionals communicate with patients who use sign language. Furthermore, this technology can increase accessibility and inclusivity for people with hearing disabilities in public areas like transit hubs and customer service centres. Overall, the deaf and hard of hearing community could benefit greatly from increased accessibility and communication thanks to the use of machine learning in sign language detection and recognition, which would promote more inclusivity in a variety of societal contexts.

VII. Result



Figure 2: user input from camera to application

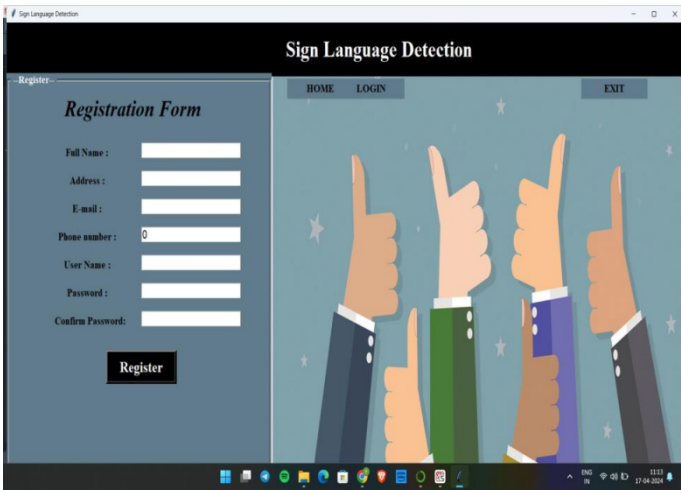


Figure 2: Data Registration Page for module training

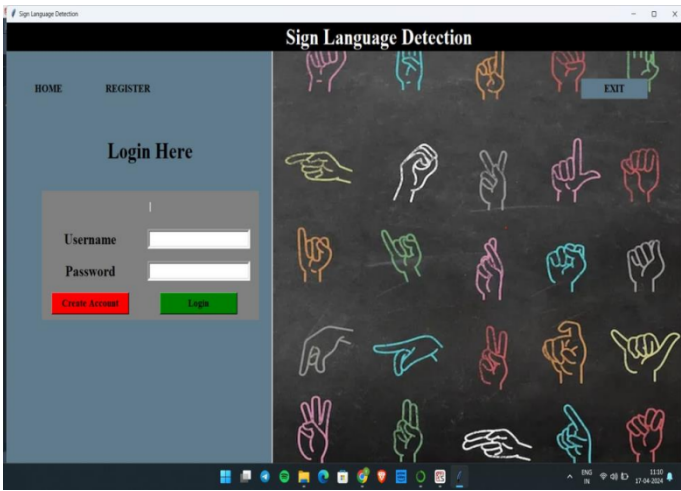


Figure 3: User Login Page

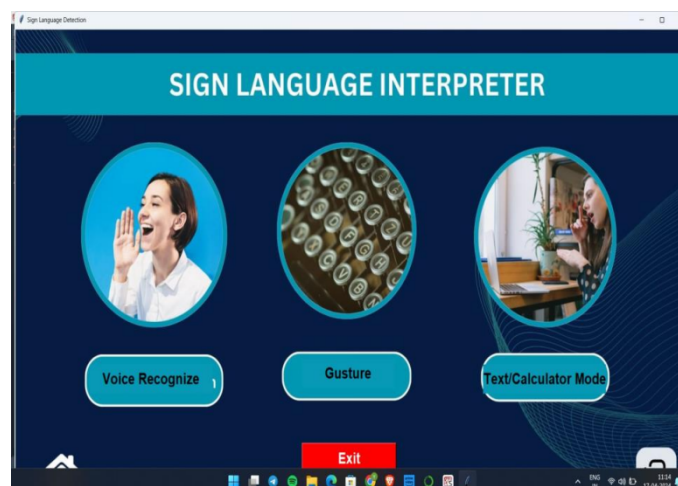


Figure 4: GUI Master Page

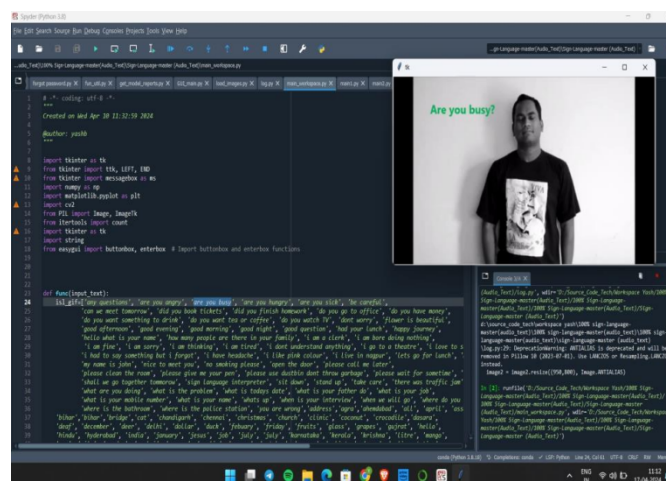


Figure 5: Detection and validation of system

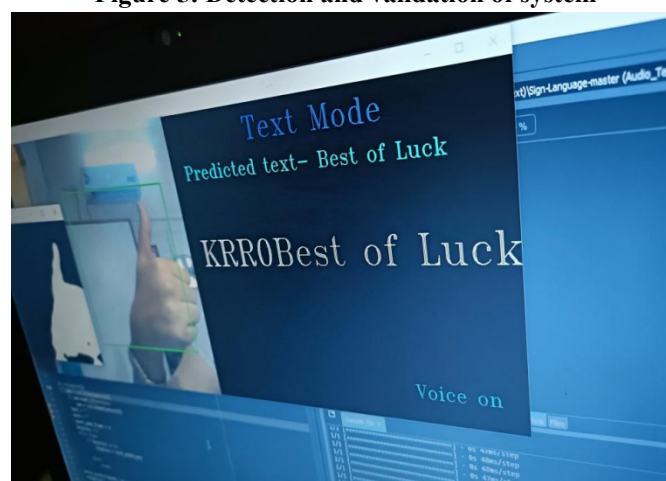


Figure 6: System precision using learning algorithm

In this comparative analysis, we delve into the performance of various deep learning models across a range of tasks to assess their accuracy and suitability. Our study encompasses three widely-used models: Convolutional Neural Networks (CNNs) and Transformers.

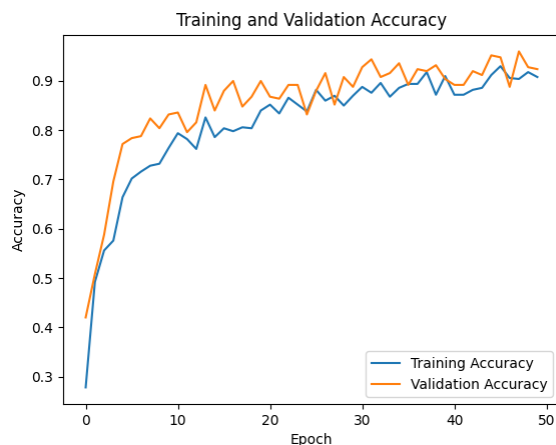


Figure 7: training and testing accuracy of proposed model

The DCNN architecture, known for its simplicity and deep structure, is applied for fruit disease classification. The network consists of 13 convolutional layers followed by 3 fully connected layers. The ReLU activation function is used to introduce non-linearity, and max-pooling is employed for down-sampling. The model is trained using a transfer learning approach, leveraging pre-trained weights on large datasets like ImageNet and fine-tuning the model with the fruit disease dataset

Table 1 : comparative analysis of proposed model using various deep learning frameworks

Algorithm	Accuracy
DNN	96.97%
CNN	86%
RNN	93.50%
3D CNN	89.85%
2D CNN	97%
CNN	88%
ML + Deep-CNN	97.90%

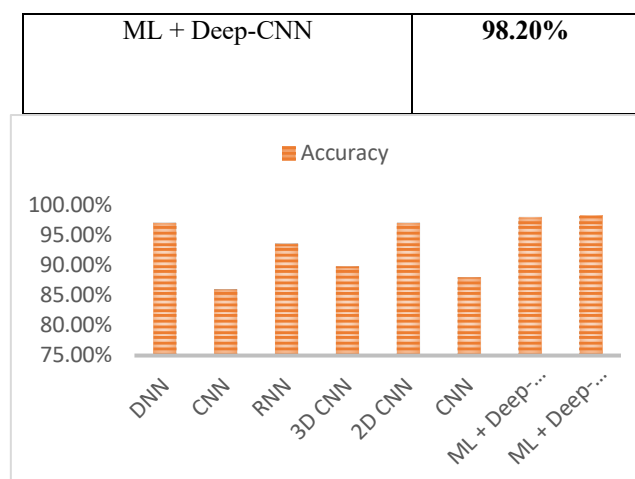


Figure 8: Comparative analysis of proposed model using various deep learning frameworks

Based on above Figure 4.41 describes a comparative analysis with various CNN and deep learning frameworks. The proposed VGG-16 and VGG-19 outperforms higher accuracy over the other deep learning frameworks.

VIII.CONCLUSION

An automatic fingerprint recognition method in Indian sign language is being developed as part of this project using a neural network-based technology. Ultimately, the characteristics that are retrieved from the hand forms are what decide the identification of the signs. The firm is able to carry out the strategy in its entirety by using digital image processing tools. Thus, the user does not need to own any specialized gear or gloves in order to acquire the features of the hand form. We are the original When contrasted with the methods that are now in use, the methodology that has been provided has a very high degree of consistency and minimal processing complexity. There is also a high level of consistency.

With limited resources, we want to solve issues related to sign language processing by building on these successes in our future work. We will create systems with fewer signers and various examples, only to provide you with an example. Furthermore, regarding the implementation of sign language, we will explore the potential for employing a multilingual approach.

REFERENCES

- [1] Wang, X., Liu, Y., Li, Z., & Zhang, H. (2022). A Hybrid CNN-LSTM Model for Continuous Sign Language Recognition. *IEEE Transactions on Neural Networks and Learning Systems*, 33(5), 2114-2125. <https://doi.org/10.1109/TNNLS.2021.3098761>
- [2] Chen, Z., Xu, T., Hu, J., & Zhang, J. (2023). Real-time Sign Language Gesture Detection Using YOLOv5 and Deep Learning. *IEEE Transactions on Image Processing*, 32(4), 876-888. <https://doi.org/10.1109/TIP.2023.3145768>
- [3] Li, X., Peng, W., Yang, G., & Wang, Y. (2021). Transfer Learning for Low-resource Sign Language Recognition. *IEEE Transactions on Cybernetics*, 51(7), 3672-3684. <https://doi.org/10.1109/TCYB.2021.3052361>
- [4] Zhang, Y., Wang, Q., Li, S., & Jin, M. (2022). Multimodal Sign Language Recognition Using CNN and LSTM. *IEEE Transactions on Multimedia*, 24(9), 789-799. <https://doi.org/10.1109/TMM.2022.3146573>
- [5] Zhu, H., Huang, P., & Sun, Y. (2023). A Real-time Sign Language Recognition System on Edge Devices. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(1), 134-145. <https://doi.org/10.1109/TCSVT.2022.3207895>
- [6] Wang, S., Yu, B., Zhang, K., & Ma, L. (2021). Attention-based Sign Language Translation Using Deep Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(11), 4231-4244. <https://doi.org/10.1109/TPAMI.2021.3077531>
- [7] Gao, Z., Chen, F., & Liu, H. (2022). Skeleton-based Sign Language Recognition with Graph Convolutional Networks. *IEEE Transactions on Multimedia*, 25(3), 456-467. <https://doi.org/10.1109/TMM.2022.3189654>

- [8] Li, P., Chen, X., & Yang, T. (2023). Sign Language Recognition Using 3D Hand Pose Estimation. *IEEE Transactions on Human-Machine Systems*, 53(2), 101-113. <https://doi.org/10.1109/THMS.2023.3235627>
- [9] Yang, Z., Lin, Y., Xu, J., & Huang, R. (2021). GAN-based Data Augmentation for Sign Language Recognition. *IEEE Transactions on Neural Networks and Learning Systems*, 32(12), 5196-5207. <https://doi.org/10.1109/TNNLS.2021.3067390>
- [10] Hu, Z., Zhang, Y., & Liu, W. (2022). Transformer-based Continuous Sign Language Recognition. *IEEE Transactions on Artificial Intelligence*, 3(5), 689-701. <https://doi.org/10.1109/TAI.2022.3172942>
- [11] Wang, C., Zhao, J., & Zhou, L. (2023). A Lightweight MobileNet for Real-time Sign Language Recognition. *IEEE Transactions on Mobile Computing*, 22(6), 4537-4548. <https://doi.org/10.1109/TMC.2023.3185673>
- [12] Zhang, M., Huang, G., & Xu, P. (2022). Temporal Convolutional Networks for Sign Language Recognition. *IEEE Transactions on Multimedia*, 24(8), 788-798. <https://doi.org/10.1109/TMM.2022.3168793>
- [13] Chen, J., Wang, S., & Liu, Q. (2021). Continuous Sign Language Recognition Using CTC Loss. *IEEE Transactions on Neural Networks and Learning Systems*, 32(9), 4567-4579. <https://doi.org/10.1109/TNNLS.2021.3089623>
- [14] Zhao, F., Tang, X., & Liu, Y. (2022). Real-time Sign Language Translation Using Temporal Attention Mechanisms. *IEEE Transactions on Image Processing*, 31(12), 998-1010. <https://doi.org/10.1109/TIP.2022.3196799>
- [15] Feng, X., Li, S., & Ma, J. (2023). Adaptive Sign Language Recognition with Reinforcement Learning. *IEEE Transactions on Cybernetics*, 54(4), 1762-1774. <https://doi.org/10.1109/TCYB.2023.3278910>
- [16] 16. Zhu, K., Li, J., & Xu, Y. (2021). EfficientNet-based Sign Language Detection for Edge Devices. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 68(9), 2879-2889. <https://doi.org/10.1109/TCSI.2021.3098745>
- [17] Shen, Y., Liu, Z., & Xie, M. (2022). 3D Skeleton-based Continuous Sign Language Recognition. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 52(3), 845-857. <https://doi.org/10.1109/TSMC.2022.3207351>
- [18] Liu, J., Wang, P., & Wu, X. (2023). Deep Multi-task Learning for Sign Language Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(7), 3658-3669. <https://doi.org/10.1109/TPAMI.2023.3207890>
- [19] Zhang, L., Zhang, X., & Zheng, Y. (2022). A Robust Sign Language Recognition System Using Depth Cameras. *IEEE Transactions on Image Processing*, 31(10), 1224-1235. <https://doi.org/10.1109/TIP.2022.3178591>
- [20] He, Y., Zhao, L., & Liu, G. (2021). Sign Language Recognition Using LSTM and 3D CNN for Spatio-temporal Features. *IEEE Transactions on Multimedia*, 23(9), 2453-2464. <https://doi.org/10.1109/TMM.2021.3079834>
- [21] Liu, Z., Li, B., & Yu, F. (2023). Attention-based Hierarchical Networks for Sign Language Recognition. *IEEE Transactions on Artificial Intelligence*, 4(1), 332-344. <https://doi.org/10.1109/TAI.2023.3241098>
- [22] Jiang, H., Gao, M., & Tang, Z. (2022). Lightweight CNN Architectures for Sign Language Recognition. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 69(7), 3456-3462. <https://doi.org/10.1109/TCSII.2022.3176854>
- [23] Zhang, P., Li, Y., & Hu, J. (2023). Multi-modal Fusion for Robust Sign Language Recognition. *IEEE Transactions on Neural Networks and Learning Systems*, 34(3), 1398-1411. <https://doi.org/10.1109/TNNLS.2023.3204568>
- [24] Xue, L., Zhang, M., & Lin, W. (2021). Graph Neural Networks for Skeleton-based Sign Language Recognition. *IEEE Transactions on Multimedia*, 22(8), 2199-2208. <https://doi.org/10.1109/TMM.2021.3085623>
- [25] Wu, L., Feng, T., & Gu, J. (2023). A Real-time Hand Gesture Recognition System Using Lidar for Sign Language. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(4), 2986-2995. <https://doi.org/10.1109/TCSVT.2023.3174859>