Hand Gesture Sign Language Recognition through Machine Learning

Dr. Mahesh Kumar Porwal
Associate Professor
Electronics and Communication Engineering
Sree Chaitanya College of Engineering,
Karimnagar, Telangana, India

Mr. Nishant Porwal
B.Tech Scholar
Department of Computer Science and Engineering
National Institute of Technology (NIT)
Delhi, India

Abstract—The research objective is to explore various ways in which machine learning models can be leveraged to detect and interpret sign language, ultimately contributing to real-time communication enhancements for the deaf community. We are excited to present the development and deployment of recognition model for sign language based on range of different models. This innovative approach not only benefits sign language learners by providing a tool for practicing their signing skills, but it also holds immense potential in bridging communication gaps. Throughout the project, we critically assessed several human- computer interface methods for gesture recognition, ultimately concluding that a combination of image processing techniques and human-hand movement classification represents the most effective approach. Notably, our system demonstrates the capa- bility to accurately recognize selected sign language signs with an impressive accuracy range of 80% to 95%, even in uncontrolled background and normal lighting conditions. This paper presents side by side comparison between two different feature based classification models - Random Forest Classifier and ANN based Classifier. The experiment is done on 2 different sign language datasets - ISL and ASL. The main task of the classification models is to recognize the characters indicated by the hand gestures of the input image. Our methods use the datasets that contains digits and alphabets applying certain machine learning techniques, thus, acquiring the accurate model and benchmarking between the models that was applied. Hence, we have given an efficient solution and result so that it can overcome and bridge the gap in the communication between the individuals with hearing and dumb disabilities.

Index Terms—Hand Gesture Recognition, Machine Learning, Image processing, American sign language, Indian Sign Language, multi-class classification, feature extraction, Artificial Neural Network (ANN), Random Forest Tree (RFT)

I. INTRODUCTION

Sign language recognition began in the early 1990s and uses rule-based systems and basic pattern-matching techniques. The field since then made significant process over the years as the advancements in computer and machine learning technologies the gesture recognition have significantly enhanced the naturalness and overall practicality of human-computer interaction. Particularly, the gesture of the characters also known as air writing has gained more attention these days by using bare hands as the gesturing object. This report briefly overviews the hand gesture translation and detection system of sign languages, discussing its significance, stages and challenges. According to a cross-sectional survey, at least around 15-20% i.e., more than 1.5 billion people of the world's population are either deaf or dumb. India the most populated country in the world is home to around 63 million people (i.e., about 6.3% of the population) who belong to the DHH (Deaf and Hard of Hearing) community. These individuals depend and rely on sign language, as an essential part of their daily lives to communicate and interact with others. The complexity and difficulty of sign languages and the lack of widespread understanding of the gestures create problems and issues in their communication with the wider world.

A. Background

Sign language is not just a communication tool but it is an important integral part of their daily lives to go hand in hand with the world. Despite their significance, it has historically faced many obstacles and issues such as misconception, lack of educational resources and limited technology. However, with the advancement in machine learning, deep learning and computer vision provide an effective solution towards the challenges and issues faced during interaction and communication. The detection and recognition process involves several steps to make the system understandable by the computers, which includes data processing, segmentation, feature extraction, pattern matching, class recognition and selecting the appropriate model. These models are generated and designed to enable real-time interaction systems between both human-to-human and human-computer, hereby improving and reducing the communication and educational gap among paired and impaired individuals. It also provides and empowers individuals with hearing and speaking impairments to be more independent and self-reliant in communicating and interacting with the world. The advancement and progress of sign language detection along with the improvement of modern technology, it has the potential to significantly improve the communication barriers in the DDH community.

II. LITERATURE REVIEW

Our literature survey has examined the behaviour of different machine learning models and algorithms that can detect and recognize sign languages with human hand gestures.

Many research studies have explored the performance and effectiveness of the models and techniques used for detection and identification of different sign languages.

In many cases, researchers have assessed how good and well the model performed based on various datasets. For example, the ASL dataset [1] which contains the images of alphabets and letters that were captured using a webcam was used to analyze the overall performance and accuracy of the model and similarly, a dataset for ISL was extracted from Kaggle [2] which contains digits from 0-9 and alphabets a-z were used and initialized to analyze and observe the overall performance and accuracy of the model that was utilized for the detection of hand-gesture. A research paper, which was performed using self-captured datasets based on HAND GESTURE RECOGNITION by USING MACHINE LEARNING models AND COMPUTER VISION algorithms [3] which can detect only a single hand i.e., with only 21 key points using MediaPipe and OpenCV libraries. The Machine learning models like the k-nearest Neighbour, Decision Tree, and Support Vector Machine were used to measure and assess the performance i.e., the accuracy, precision, and recall of the model in the detection process which gave insight into the limitations and effects of the model. An image-based Indian sign language recognition using deep neural networks was performed by Harleen Kaur, Sanjam Kaur Bedi and M A. Lekhana [4] which was focused on generating the maximum dataset of ISL and designing the best model for CNN (Convolutional Neural Network) to train and process the image and tried to achieve the maximum accuracy of the model. It used AlexNet, InceptionV3, Restnet-50, and VGG-16, which gave the overall performance of different models, it has allowed us to look at and understand the in-depth models and algorithm for the detection of sign languages. Feature extraction is also a major issue in the detection of sign language for better view and understanding a research paper based on feature extraction: a survey of types and techniques by Ayodejii and Shruti Jain [5] was performed where they explained the importance of extracting related information from raw data and also the features for extraction of data and to implement effectively. There are different types of features which are classified as general features (GF) and domain-specific features (DSF) which include such as local, global and pixel-level features for analysing the image. Here, colour features play an important role in the extraction of visual features. It includes colour spaces like HSV, LUV and the most commonly used RGB which helps to define the image features. It has used the Gabor filter to extract the pattern and homogeneity of the image extraction. This paper also classifies face feature extraction which involves steps such as detection, classification, alignment and representation for accurate detection and identification.

Researchers have explored sign recognition in a deep survey, Rastgoo presented a journal paper on Sign Recognition: A deep survey [6], where he describes how sign language recognition is divided into 2 categories: standalone sign language identification and continual sign language recognition. It also gives sign language recognition areas divided into face sign language, human sign language and most commonly, hand sign language. This paper basically provides a comprehensive review of how can we detect sign language using different types of models and methods of deep learning such as CNN and RGB-based human detection methods. It also describes the challenges and issues that have been faced during the extraction of datasets, prepossessing of the image, extraction of features and choosing the appropriate models and techniques. As with the advancement of technology, many researchers have come up with ideas and significantly contributed in the detection of sign language through different methods and techniques. A paper entitled: Recent Development in Visual Sign Language Recognition [7], describes a comprehensive approach to the concept of visual sign language recognition, highlighting the importance of both manual and facial features for accurate and precise recognition and detection also it addresses the challenges and difficulties of extraction and processing of features to achieve the robust and accurate to improve the real-time use of sign language detection system [8]. Furthermore, the researchers have contributed and performed for the development of more accurate and powerful models and techniques for feature extraction and detecting sign languages more accurately. Convolution neural networks (CNN) have been utilised by Lionel Pigou, Sander Dieleman, Pieter-Jan Kindermans, and Benjamin Schrauwen for recognising sign language [9], gives insights on sign language recognition and detection using the CNN models, through this model they achieved training accuracy of 91.7%. the paper mainly focuses to detect and recognise sign languages automatically. It uses Microsoft Kinect, CNN, and GPU which includes advanced machine learning models and algorithms, high-performance computing software and hardware and as well as depth-sensing cameras, which helped them to create a system that can facilitate conversation and interpretations in the deaf community and has insured to help in the detection of sign language in real-time. Through the help of all that technology and models, these researchers have tried to obtain an accurate and more efficient method. American sign language (ASL) detection system was developed in the United States by a group of researchers team [10], which was focused on only 30 words. The research involves tracking hand movement, it uses the Markov model (HMM) for classification. They tested the system using the RWTH-BOSTON-50 dataset [11], where their system acquired an accuracy of 89% with an 11% error rate. Dong Ping Tian performed research on techniques for capturing and modelling image data [12], which focuses on the importance and representation of the images, it also describes the challenges of extracting the features from the image, through this research he tried to fill up the gap of the previous studies, research that was failed to cover and also gives a vision about the latest technology that has been developed.

In conclusion, through this literature review, has provided and given insights and revealed the diverse deep learning and machine learning models and techniques that are applicable

during image extraction, image processing and specifically in the creation of the system of sign language recognition with the ability to choose the right model and use the best techniques, acquiring the accurate result.

III. SPECIFICATION AND DIFFERENCES OF ISL AND ASL

Sign language has been the major factor for the people with deaf and hard-hearing people, In the 1500s, a Spanish Benedictine monk Pedro Ponce de Leon, who was also considered the first teacher and the first person to develop a method for teaching the deaf community. In 1620 another Spanish Priest and linguist Juan Pablo Bonet first published the first manual alphabet system for the deaf community. Sign language is a primary part of the deaf and hard-hearing community globally. It is a natural language but a complex system for communication and interaction. It has its own grammar, syntax, vocabulary and distinct gesture or pantomime. It has a higher degree of iconicity but it varies across different languages and concepts. Now, many countries have legalized and recognize their sign languages, as sign languages are not universal, different regions or countries have their own unique sign languages. Sign languages rely on the visual-gestural forms of communication rather than vocal channels, it is perceived through hand gestures, movement, expression and body postures. Some components of sign languages are palm orientation, shape, movement and non-manual markers. It has a unique grammatical structure that is different from the normal languages, as it uses a topic-comment structure rather than a subject-verb-object structure. As with technological advancement in machine learning and computer vision, many researchers have contributed and emerged to develop a system that will be able to recognize, translate, detect and interpret sign languages easily, so that it can be utilized and implemented in every part of the world and also support the deaf and hard- hearing community. There are two very popular sign languages, the ASL and ISL also known as the American Sign Language and Indian Sign Language, each have their own differences according to the region and environment. ASL also known as American Sign Language, is one of the most popular and highly used sign languages around the world, it was developed in the early 19th century in the United States and is widely used in the US and some parts of Canada. It was inspired by French sign language and indigenous sign languages used by Native American communities. It has its grammar and syntax and utilizes only a one-handed system. Some of the key components of 8ASL are handshapes, location, movement, palm orientation, and non-manual markers. It uses a three-dimensional space, spatial grammar, classifier and fingerspelling. As like all other sign languages ASL has its sociolinguistic variations such as age, ethnicity and educational background. In American society, ASL plays a major role in the Deaf and mute community. It typically follows SOV (Subject-Object-Verb) word and has 26 distinct alphabet handshapes, where the handshapes are used to represent the movement and describe its size and shape, ASL is complex but efficient natural sign language. As with the advances in

Fig. 1. American Sign Language

technology mainly in machine learning, computer vision and other types of equipment, researchers have been putting their effort and have come up with many techniques, models and ideas for easier ways of utilisation, learning and implementation, thus supporting the deaf and hard- hearing community across the world. Researchers such as Scott Liddell, helped to understand the use of ASL grammar, William Stroke, in 1960 founded the linguistic study of ASL and established the language with its grammar. Karen Emmorey also made an experiment by comparing brain processing sign language with normal spoken language. As with the increase in research on ASL, many researchers have improvised and plotted new ideas and provided a gateway for the development of better models and techniques for the interpretation of American Sign Language. A team led by researcher, Helen Cooper, using the RNN (Recurrent Neural Network) and LSTM (Long Short-Term Memory) models developed a system that can detect the ASL gesture. Another researcher Alexei Yankovich and Jian Wang used CNN for extracting features and applied deep learning techniques to detect ASL from a video input. Researchers like Lionel Pigou and his team worked on the interpretation and detection of full sentences based on ASL datasets. These researchers and teams have pushed and provided their best to revolutionize communication tools for the deaf and mute community. As with the development in technology, many researchers and teams are coming up and pushing the boundaries in the field of American sign language detection systems.

The early 20th century marked the beginning of formal deaf education in India; however, sign languages were not practised and were not given importance rather than oral methods were initially focused and taught. The mid-20th century was the period when people recognised the importance of sign language and saw the efforts and desire to study and document sign language across India the 1980s and 1990s, were the years when India tried and attempted to standardize and implement Indian Sign Language however in 2016 marked the year when the Government of India officially recognize Indian Sign Language (ISL) as a language in its own right, which was considered a breakthrough for the community of

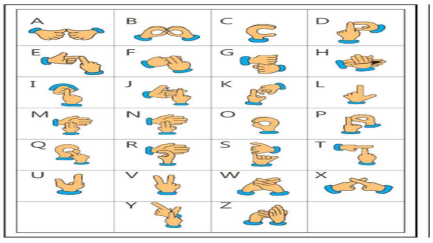

Fig. 2. Indian Sign Language

deaf and hard- hearing in India. In recent years we have seen the development of ISL dictionaries, teaching material and other programs that help and support the deaf and mute community in India. As with the increase in the number of researchers, it has begun to study, research and experiment beyond the boundaries. Many researchers like Ankit Ojha and Dipti Prakash Mohapatra have used CNNs, deep learning techniques to detect and interpret sign language through gestures from image and video data as input. Another researcher Kumud Tripathi used some techniques like Histogram of Oriented Gradients (HOG) and Scale-invariant Feature Transform (SIFT) with machine learning models and classifiers to extract the image feature aimed to make the technology more accessible and accurate translation. Many other researchers have contributed and tried in the development of ISL detection systems as ISL detection systems using machine learning techniques, models and classifiers is still an emerging field, unlike ASL which has extensive research and results. Some of its key components which include handshapes, two-handed signs, non-manual features, spatial grammar, movement and palm orientations. As ISL uses a double-handed sign system there are many challenges faced in the creation of ISL detection systems, some of the main challenges that are faced in ISL detection systems are feature and image extraction, limited availability of a large number of ISL datasets, real- time processing, continuous sign recognition, linguistic complexity along with the standardisation and most importantly balancing accuracy and implementation across different environments. Despite many challenges and issues researchers are making an effort and pushing beyond their limits in the development and progress in improving the ISL detection system and trying to provide a better technological system to the people of India as well as worldwide.

IV. APPROACH

The recognition of the Sign Language is done on 2 steps. The first step is extracting the feature map of the image and then mapping the features with respective letter and digits. The feature extraction process is done by using google's mediapipe library. In the second step, the extracted features aka landmarks are sent to the classifier as input and the classifier recognize the digit or letter indicated by the sign language. To properly use the classifier model, it has to be trained first. So the dataset is split in 80:20 ratio for training and validation. Random Forest Classifier comprises of a group of decision tree and then the output is retrieved by applying an ensemble method. Artificial Neural Network is a Machine

American Sign Language (ASL)	Indian Sign Language (ISL)
It was developed in the early 19th century in the USA	ISL is a naturally founded and evolved language in India.
It is used and known most popularly in the USA and some parts of Canada	It is used and known only in India
ASL is inspired by French Sign Language	ISL is inspired by Indigenous sign system of India
ASL is a single or one-handed system of communication.	ISL is a double or two-handed system of communication.
ASL is based on the Latin alphabet	ISL is based on Devanagari Script
It uses the SOV (Subject-Object-Verb) structure	It generally uses an SOV structure
ASL has numerous researchers and resources	ISL is a growing and developing system with a limited number of researchers and resources
It follows and reflects the American cultural context	It follows the diverse Indian cultural elements
American Sign Language is a highly standardized and well-established sign language system.	Indian Sign Language is a regional variation with less standardization and a growing system community.

TABLE I
THE KEY DIFFERENCES BETWEEN AMERICAN SIGN LANGUAGE (ASL) AND INDIAN SIGN LANGUAGE (ISL)

Learning model uses a network of perceptrons which can be trained and optimized to solve a problem of specific area depending on the provided training data and the model's capacity.

V. RESULT

In this section, we give a brief description of the results and analysis that have been obtained during the research by implementing and using different types of machine learning models and algorithms. Further in the section, the models are been benchmarked, resulting in the best and most accurate models that can be used in the detection of Indian sign language using hand gestures. The dataset that was used for the operation was collected from Kaggle and it contains all 10 digits (0-9) and 26 alphabets (a-z) captured in image form, it was extracted for both ASL and ISL. The link is cited below for future reference.

A. American Sign Language (ASL)

Dataset Source: https://www.kaggle.com/datasets/ayuraj/asl-dataset

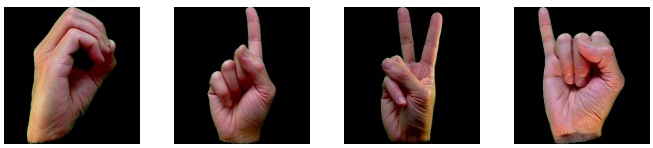

Fig. 3. Some dataset samples used for training ASL based Model

1) Random Forest Classifier: The result from the given inputs as in Fig. 3 are given in Fig. 4.

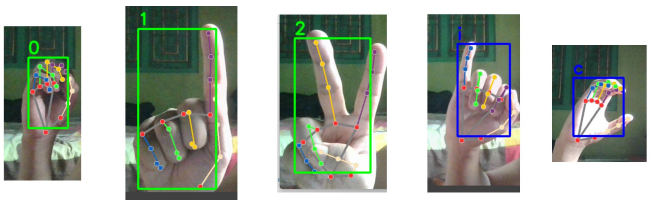

Fig. 4. ASL recognition with Random Forest Classifier

Model Structure

The structure of the model that was implemented is described below:

- Input Feature Size: 42
- Output Label Classes: 36
- The number of Decision Trees in the Random Forest Model is 100

Training Parameters

The accuracy of the training data along with the size of the training datasets is presented as follows:

- The Model has 100.00% Accuracy for Training Data
- Sample Size of the training dataset is 1471

Evaluation Metrics

The accuracy of the testing data along with the sample size that was evaluated is presented as follows:

- The Model has 98.64% Accuracy for Test Data
- The sample size for the Evaluation dataset is 368

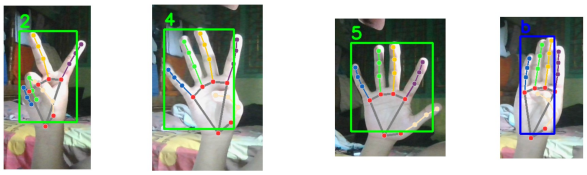

Fig. 5. ASL recognition with ANN

2) Artificial Neural Network: The result from ANN by the given inputs as in Fig. 3 are presented in Fig. 5. The structure of the model along with the optimizer and loss function used is presented as follows:

Layer (type)	Output Shape	Param #
Dense	(None,36)	1,548

- Total params: 1,548 (6.05 KB)
- Trainable params: 1,548 (6.05 KB)
- Non-trainable params: 0 (0.00 B)
- Loss Function Used: Mean Squared Error
- Optimizer Used: AdamW with a learning rate of 0.01

Training Parameters

- The accuracy of the model on training data is 93.43%
- Count of Training data: 1461

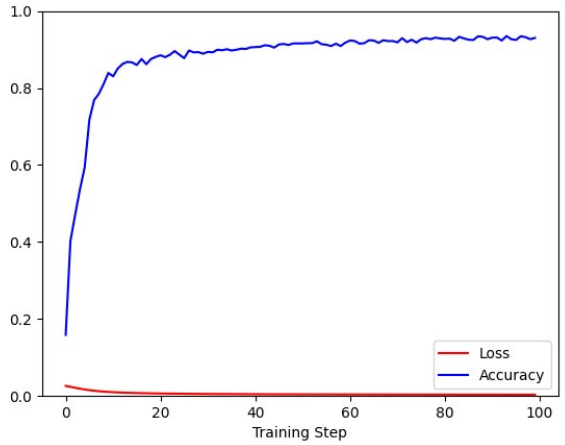

Fig. 6. Training parameters of ANN for ASL

Evaluation Metrics

- The accuracy of the model on test data is 81.48%
- Count of Test data: 378

B. Indian Sign Language (ISL)

Dataset Source:
https://www.kaggle.com/datasets/prathumarikeri/indian-sign-language-isl

Fig. 7. Some dataset samples used for training ISL based Model

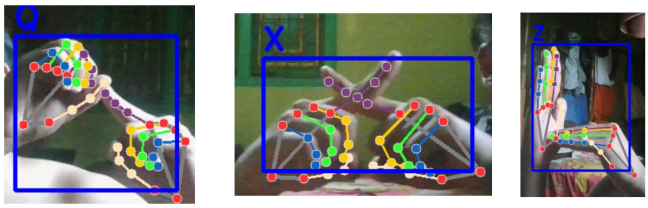

Fig. 8. ISL recognition with Random Forest Classifier

1) Random Forest Classifier: The result from the given inputs as in Fig. 7 are given in Fig. 8.

Model Structure

The structure of the model that was implemented is described below:

- Input Feature Size: 84
- Output Label Classes: 36
- The number of Decision Trees in the Random Forest Model is 100

Training Parameters

The accuracy of the training data along with the size of the training datasets is presented as follows:

- The Model has 100.00% Accuracy for Training Data
- Sample Size of the training dataset is 33905

Evaluation Metrics

The accuracy of the testing data along with the sample size that was evaluated is presented as follows:

- The Model has 99.88% Accuracy for Test Data
- The sample size for the Evaluation dataset is 8480

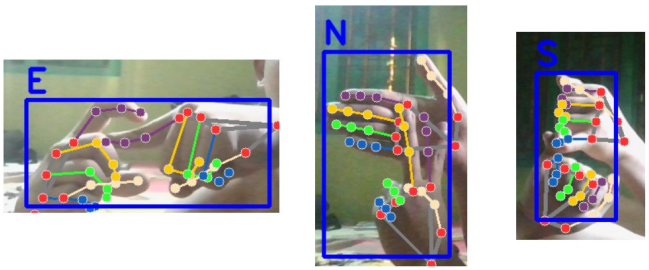

Fig. 9. ISL recognition with ANN

2) Artificial Neural Network: The result from ANN by the given inputs as in Fig. 7 are presented in Fig. 9. The structure of the model along with the optimizer and loss function used is presented as follows:

Layer (type)	Output Shape	Param #
Dense	(None,36)	3,060

- Total params: 3,060 (11.95 KB)
- Trainable params: 3,060 (11.95 KB)
- Non-trainable params: 0 (0.00 B)
- Loss Function Used: Categorical Crossentropy
- Optimizer Used: AdamW with a learning rate of 0.01

Training Parameters

- The accuracy of the model on training data is 25.42%
- Count of Training data: 33905

Evaluation Metrics

- The accuracy of the model on test data is 25.46%
- Count of Test data: 8480

C. Model Benchmarking

The complete comparison of the model with the accuracy of testing and training models along with the classifiers implemented, which describes the differences and shows the most accurate result of the research is presented as follows.

Models	Training Accuracy	Testing Accuracy	Testing Samples	Training Samples
Random Forest	100.00%	98.64%	368	1471
ANN	93.43%	81.48%	378	1461

TABLE II: Model Benchmarking for ASL

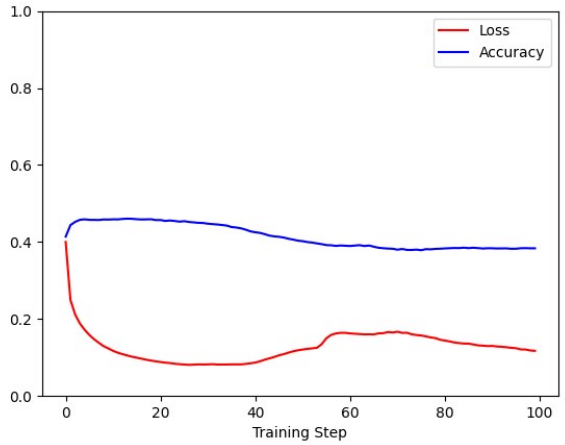

Fig. 10. Training parameters of ANN for ISL

Models	Training Accuracy	Testing Accuracy	Testing Samples	Training Samples
Random Forest	100.00%	99.88%	8480	33905
ANN	25.42%	25.46%	8480	33905

TABLE III: Model Benchmarking for ISL

VI. Conclusion

The results presented so far procures that the Random Forest Model performs better than the simple Artificial Neural Networks for both ISL and ASL datasets. In Simple ANN, the Model that used Mean Squared Error as the loss function while training performed better than the one that used categorical cross-entropy. For this experiment, it is found that the Random Forest Classifier model performs better than the Artificial Neural Network based classifier.

VII. Future Scope

Collecting and making the sign language dataset larger and with multiple sign languages in order to make them more reliable and provide accurate results in any environment.

VIII. References

[1] "Asl dataset," https://www.kaggle.com/datasets/ayuraj/asl-dataset.
[2] "Indian sign language (isl) dataset," https://www.kaggle.com/datasets/prathumarikeri/indian-sign- language-isl.
[3] A. Singh, A. Wadhawan, M. Rakhra, U. Mittal, A. Al Ahdal, and S. Jha, "Indian sign language recognition system for dynamic signs," 10 2022, pp. 1–6.
[4] A. Salau and S. Jain, "Feature extraction: A survey of the types, techniques, applications," 03 2019, pp. 158–164.
[5] R. Rastgoo, K. Kiani, and S. Escalera, "Sign language recognition: A deep survey," Expert Systems with Applications, vol. 164, p. 113794, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S095741742030614X
[6] U. von Agris, J. Zieren, U. Canzler, B. Bauer, and K.-F. Kraiss, "Recent developments in visual sign language recognition," Universal Access in the Information Society, vol. 6, no. 4, pp. 323–362, Feb 2008. [Online]. Available: https://doi.org/10.1007/s10209-007-0104-x
[7] A. M. M. Y. Kaustubh Jadhav, Abhishek Jaiswal, "Sign language recognition using neural network," Journal of Electronics and Telecommunication Engineering, vol. 1, no. 1, pp. 1–12, 2024,
[8] L. Pigou, S. Dieleman, P.-J. Kindermans, and B. Schrauwen, "Sign language recognition using convolutional neural networks," in Computer Vision - ECCV 2014 Workshops, L. Agapito, M. M. Bronstein, and C. Rother, Eds. Cham: Springer International Publishing, 2015, pp. 572–578

.

[9] B. Garcia, S. A. Viesca et al., "Real-time american sign language recognition with convolutional neural networks," Convolutional Neural Networks for Visual Recognition, vol. 2, no. 225-232, p. 8, 2016.

[10] M. Zahedi, D. Keysers, T. Deselaers, and H. Ney, "Combination of tangent distance and an image distortion model for appearance-based sign language recognition," in Deutsche Arbeitsgemeinschaft für Mustererkennung Symposium, ser. Lecture Notes in Computer Science, vol. 3663, Vienna, Austria, Aug. 2005, pp. 401–408.

[11] D. Tian, "A review on image feature extraction and representation techniques," International Journal of Multimedia and Ubiquitous Engineering, vol. 8, pp. 385–395, 01 2013.