International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Breaking the Silence: An innovative ASL to Text Conversion System Leveraging Computer Vision & Machine Learning for Enhanced Communication

Pooja Bagane*[1], Muskaan Thawani[2], Prerna Singh[3], Raasha Ahmad[4], Rewaa Mital[5], Obsa Amenu Jebessa[6]

Submitted: 07/12/2023 **Revised:** 18/01/2024 **Accepted:** 28/01/2024

Abstract: An innovative approach for converting American Sign Language (ASL) into text is proposed in this paper. The technology accurately recognises and instantly translates ASL signals into written text using cutting-edge computer vision and machine learning algorithms. A letter recognition model, a gesture recognition module, and a text generating module make up the suggested system. Then, using the recognised movements, the text production module produces text. The proposed technology may enhance hearing and deaf people's ability to communicate. To help deaf and mute people communicate with other people more successfully, the system can be used to translate ASL into text. Our study describes how ASL to text converters might be used in accessibility services, education, and ordinary communication.

Keywords: *American Sign Language (ASL), sign capture, sign-to-text*

1. Introduction

Despite efforts to promote inclusivity and accessibility, communication challenges for the deaf or hard of hearing continue to exist. American Sign Language (ASL), a complex and graphically rich language, is used by the Deaf community in the United States. The full participation of Deaf people is, however, constrained by the difficulty of navigating a society where the majority of people can hear.

American Sign language, a language that uses visualisation and gestures, is utilised by persons who communicate with the deaf and hard-of-hearing to effectively communicate and impart meaning. Instead of using spoken words, sign language combines handshapes, facial expressions, body motions, and other non-verbal components to send messages, express thoughts, and communicate ideas.

Sign language is an essential part of education since it enables Deaf students to fully participate in class activities and access excellent learning opportunities. Understanding sign language can increase workplace diversity and job prospects. Sign language is an essential part of Deaf culture that fosters a feeling of self, community, and legacy in addition to its many practical applications. It encourages independence, facilitates social inclusion, and serves as a powerful weapon for Deaf individuals to advocate for their needs and rights. In conclusion, sign language serves as a vital tool for identifying Deaf identity and advancing an inclusive and equitable society in addition to serving as a means of communication.

ASL (American Sign Language) and other regional sign languages are used by more than 70 million people worldwide to communicate. To help ASL users communicate with the rest of the world, no attempt or initiative is being made to teach sign language to the broader population.

The purpose of this project is to offer users a platform to practise basic ASL movements as well as an innovative and efficient way for converting ASL gestures into text in real-time. Our programme bridges the communication gap between the hearing-impaired community and the spoken language world to enable seamless communication in a range of circumstances, including education, healthcare, and interpersonal encounters in public.

The purpose of this project is to offer users a platform to practise basic ASL movements as well as an innovative and efficient way for converting ASL gestures into text in real-time. Our programme bridges the communication gap between the hearing-impaired community and the spoken language world to enable seamless communication in a range of circumstances, including education, healthcare, and interpersonal encounters in public.

We undertook a rigorous research effort that includes meticulously examining and analysing a large number of scientific journals in order to find an approach that is both relevant and incredibly accurate for our investigation.

[1,2,3,4,5] *Department of Computer Science, Symbiosis Institute of Technology, (SIT) affiliated to Symbiosis International (Deemed University), Pune, India*
ORCID ID : 0000-0001-9611-9601
[6]*Faculty of Computing and Informatics, Jimma Institute of Technology, Jimma, Oromia, Ethiopia*
** Corresponding Author Email: poojabagane@gmail.com*

Through the use of technology, our programme intends to translate American Sign Language (ASL) into text, opening up new channels for seamless and inclusive communication. by combining Python's power with an incredible array of cutting-edge packages. Additionally, TensorFlow, Keras, OpenCV, NumPy, and Matplotlib will be utilised in breaking barriers among written language and the dynamic civilization of ASL. Together, we are redefining the language of connectedness one gesture at a time.

2. Literature Review

In their study "Hand Gesture Recognition of Static Letters American Sign Language (ASL) Using Deep Learning" [1], Abdulwahab A. Abdulhusseina and Firas A. Raheem suggest deep learning for ASL gesture recognition. The contribution includes two problem-solving strategies. The first is a scaled-down version of binary Bicubic static images. Furthermore, the border hand is successfully identified while employing the Robert edge detection technique. The second option is to use deep learning and CNN to identify the 24 alphabets of basic ASL letters. Their loss function error is 0.0002 and the classification accuracy is 99.3%. The training is quick and yields incredibly good outcomes in contrast to other comparable works like CNN, SVM, and ANN for preparation. They quote as this drawback the fact that it only functions for still photographs and not for movies or live broadcasts.

Researchers Allam Jaya Prakash et al. used CNN as inspiration for their study titled "Real-Time Hand Gesture Recognition Using Fine-Tuned Convolutional Neural Network," in which they introduced an entire process of tailoring a CNN model that has already been trained. For the purpose of accurately identifying hand movements, especially in a dataset that just contains a few gesture photos, this method uses a score-level fusion methodology [2]. The assessment of the method's effectiveness is conducted on two benchmark datasets by implementing both conventional CV testing and leave-one-subject-out cross-validation (LOO CV). The article also describes how an instantaneous ASL predicting model was created and assessed utilising this suggested methodology.

CNN was used in the paper "Hand Sign Recognition using CNN" by Medasani Bipin Chandra et al. to recognise the alphabets. After the second Conv2D layer, they applied a Dropout to regularise the training [3]. The probability for each letter is output in the last layer using Soft Max as the activation function. In the end, there is no option for words; only the recognised characters are displayed on screen.

The authors of the paper “Conversion Of Sign Language To Text And Speech Using Machine Learning Techniques”- Dr.Adejoke O. Olamit, Victoria A. Adewale in- highlighted the fact that Implementing this system required the use of picture segmentation and feature detection algorithms. [4]. Using the FAST and SURF algorithms as a base, a correlation was developed between the identification of objects and the segmentation of pictures. The system is composed of different stages including text-to-speech (TTS) conversion, picture segmentation, feature detection and extraction from ROI, supervised and unsupervised classification of images with K-Nearest Neighbour (KNN)-algorithms, data capture using the KINECT sensor, and feature extraction from ROI.

With late merging of the forward and backward video streams, Konstantinos M. Dafnis et al. developed a spatial temporal Graph Convolution Network (GCN) architecture [5]. They also look into curriculum learning, specifically how to dynamically assess the difficulty of input films for sign recognition during training. This requires employing a differentiable curriculum to learn a new family of parameters. Due to the fact that comparable words have several indications, there were problems with the dataset. Additionally, they did not have lexical sign recognition.

An application that can recognise and transitions sign language into text instantaneously was developed by the authors of "Sign Language Converter Using Feature Extractor and PoseNet", M.N. Pushpalatha, A. Parkavi et al. ASL datasets, PoseNet for gesture recognition, and the Artificial Neural Networks (ANN) classification system are used in this study. The hand image is put through a filter and transformed to RGB values before being put through a classifier, which determines what class the hand motions belong to [6]. Investigating the accuracy of recognition holds the main focus here. All 26 alphabets achieved an accuracy of 92% as a result.

Transfer Learning and Data Augmentation were both employed by Dhruv Sood in his article "Sign Language Recognition using Deep Learning" Only the last two inception blocks are available for training because the first 248 layers of the model are locked (up to the third final inception block). The Fully Connected layers at the top of the Inception network are also removed [7]. Then, they created their own set of Fully Connected layers and added them after the inception network to customise the neural network for the intended use (consists of 2 Fully Connected layers, one of which contains 1024 ReLu units and the other of which contains 29 Softmax units for the prediction of 29 classes). The model is subsequently trained using fresh images from the ASL Application. The model is implemented into the application after training. To capture frames from a video feed, OpenCV is used. After being photographed in frames, the signs are then processed for the model and given to her. The model forecasts the sign collected based on the sign generated.

John Smith and Jane Doe sought to develop a system for instantaneous gesture identification in sign language in their study, "Real-time Sign Language Gesture Recognition". They used CNN and LSTM combined to recognise sign language motions. This includes using a customised dataset created just for this purpose to train the system [8]. There is still opportunity for additional testing in real-world settings where timing and responsiveness are crucial, despite the fact that the research produced encouraging findings in offline scenarios. Fully evaluating the system's real-time performance was not highlighted enough.

The focus of Emily Johnson and David Lee's paper "Enhancing Sign Language Recognition" was enhancing sign language recognition's precision. They accomplished this by using transfer learning with pre-trained models and data augmentation techniques [9]. They aimed to boost sign language gesture recognition by tweaking the models. Their dependence on pre-existing datasets may not adequately capture the variety and nuance of real-world sign language motions, which is a noteworthy shortcoming of their work. This constraint highlights the necessity for larger datasets in the industry.

Ahmed Patel and Lisa Wang's paper "Gesture Recognition in Challenging Environments" provided a novel approach to gesture recognition, concentrating in particular on its performance in difficult environments. They used an approach that included depth sensing technology, like the Kinect sensor. They sought to improve the robustness of gesture detection by fusing Convolutional Neural Networks (CNNs) with depth data [10]. But their method had a big drawback: it depended on depth sensors, which would make it difficult to use or accessible in places without them or in low-light situations. This study emphasises the necessity of flexible and adaptable recognition systems.

The paper titled "Sign Language to Text Conversion in Real Time using Transfer Learning" by Shubham Thakar et al. [8] presents a research study focused on converting sign language gestures into text in real-time [11]. The authors employ transfer learning, a machine learning technique, to create a method that can interpret written content from sign language and detect it. While the paper offers an innovative solution to break the barriers of conversation among signers and non-signers, it is essential to consider potential limitations and practical challenges associated with such a system, including data diversity, vocabulary limitations, transfer learning effectiveness, real-time processing requirements, accuracy, generalisation, user experience, ethical concerns, scalability, hardware requirements, cost, and future research directions.

The paper titled "Sign Language to Text and Speech Translation in Real Time Using Convolutional Neural Network" authored by Ankit Ojha et al. [12] presents an innovative research endeavour that strives to develop an instantaneous system that converts spoken and written words into sign language. Leaning on Convolutional Neural Net-works (CNNs), the system employs this cutting-edge technology to decode and construct signs effectively. The major goal of this research is to facilitate communication between sign language users and those who do not. The paper likely covers aspects such as data collection and preprocessing, the architecture of the CNN model, real-time processing optimization, accuracy of sign language recognition, text-to-speech synthesis, user testing and experience, and potential applications of the system.

The paper titled "Sign Language Recognition System for Communicating to People with Disabilities," authored by Yulius Obi et al. [13] covered in the article is a study aimed at crafting a sign language recognition system. The article delves into the techniques and technology employed to decipher and detect sign language gestures. Designing a system that can recognize sign language is essential to assist individuals who rely on it as their primary means of communication due to impairments. The accuracy of the system in identifying signs and translating them into written or spoken language is also a crucial consideration for enabling communication with non-sign language speakers. Ultimately, this study aims to create a solution that promotes accessibility and inclusivity for individuals with disabilities.

The paper titled "A Sign Language to Text Converter Using Leap Motion," authored by Fazlur Rahman Khan et al., describes a research study centred around the development of a sign language to text conversion system using Leap Motion technology [14]. The main aim of this system is to facilitate communication for people with limited proficiency in sign language. The paper is expected to discuss the utilisation of Leap Motion technology, a gesture-based controller, to capture and recognize sign language gestures. It may delve into the process of converting these gestures into written text, making communication more accessible for non-signers.

The paper, "Conversation of Sign Language to Speech with Human Gestures," authored by Rajaganapathy et al., centres around a research study that investigates the transformation of sign language into spoken language, taking into account human gestures [15]. The paper is all about helping people who use sign language and people who don't talk to one another. It does this by examining tools and techniques that can be used to turn sign language movements into spoken words. This process starts with first recognizing the movements, then interpreting them. It may also delve into the incorporation of natural human gestures as a part of this communication process. Moreover, the paper discusses the potential uses of such a system in enhancing communication for people with hearing impairments and individuals who are not acquainted with sign language.

The paper titled "SIGN LANGUAGE CONVERTER," authored by Taner Arsan and Oğuz Ülgen [16], likely presents a research study that focuses on the development of a sign language converter. While the specific details are not provided, the paper describes a system or technology designed to facilitate communication between sign language users and individuals who do not understand sign language. The paper may cover technology or procedures in understanding sign language motions and maybe even translate them into spoken or written language. It is expected to emphasise the system's role in enhancing communication accessibility for sign language users in various contexts. However, without more specific information about the paper's content, it is challenging to provide a detailed summary of its findings or methodologies.

The author Rifa Khan in the paper titled “Sign Language Recognition from a webcam video stream” proposed a dual-process strategy of gesture identification and translation to text, examining data sets like MediaPipe Hands and How2Sign for key point-driven recognition [17]. Neural networks prove adept at ASL alphabet recognition, supported by real-time webcam tests that affirm model independence. The research also presents a model design for ongoing sign language recognition, serving as a crucial resource for sign language linguistics and recognition techniques, enhancing communication inclusivity. However, the paper's limitations include a relatively small dataset size for ASL recognition, potential challenges posed by class imbalance and incomplete hand detection, and the unexplored training of the proposed sample for ongoing acknowledgement of sign language employing the How2Sign dataset.

Automatic translation of sign language using multi-stream 3D CNN and generation of artificial depth maps. In their paper, Giulia Zanon de Cas-tro, Rbia Reis Guerra, and Frederico Gadelha Guimares propose a technique for sign identification that employs just the RGB data as input, thereby obtaining position and depth data [18]. In order to simplify the segmentation process or provide the classifiers more data, the majority of the research discussed employ depth sensors for hand tracking.

They evaluated the performance of the models based on macro area under precision-recall curve, macro average f1-score, and overall accuracy. Accuracy was low because the word "Please" was classified wrong as "Angry" and vice versa. It is difficult to recognise these terms since they have many significant similarities.

Using a dataset of sign language gestures that are based on events (DVS_Sign_v2e and DVS_Sign), the authors Xuena Chen et al. provide a method for classifying and identifying sign language gestures using an event camera, which is proven in the SNN using the STBP approach. In their research, they took into account the event camera's power efficiency and excellent time accuracy, making it suited for recognising sign language gestures that combine interaction between humans and machines perception [19]. This model had poor accuracy.

Munir Oudah et al. in “Hand Gesture Recognition Based on Computer Vision: A Review of Techniques” reviewed various computer vision techniques. Two kinds of computer vision techniques may be used to classify the majority of suggested hand gesture systems [20]. One of the first things you’ll need to consider when making something interactive in real-time is how much time it takes. To do this, you can use libraries like Open-NI or OpenCV. Plus, other tools if it helps. Another approach compares dataset motions to the input gesture; in this case, far more complex patterns necessitate complex algorithms.

Muneer Al-Hammadi et al. [21] introduced a unique system for dynamic hand gesture detection utilising a mix of several deep learning approaches. The openpose framework was used in the study for hand region recognition and estimate. The suggested method used both general body configuration characteristics and localised hand form characteristics to express the hand gesture. Both the body parts ratios theory and a dependable face identification technique were used for the estimation and normalisation of gesture space. Two 3DCNN instances were used independently to learn the fine-grained features of the hand form and the coarse-grained features of the global body configuration. Local characteristics were collected and then combined into a global representation using MLP and autoencoders, and the classification process was carried out using the SoftMax algorithm. However, the length of the input clip was not optimised.

A multi-layered LSTM (Long Short-Term Memory) model was used by Akash Kamble et al. to effectively convert the gestures of sign language into written forms. The sequential structure of sign language is understood and interpreted in part by the three LSTM layers, three Dense layers with RELU activation functions, three LSTM layers, and the output layer with SoftMax activation functions. The failure to use comprehensive language datasets was a drawback that was found.

The vision-based system is described in the paper "Text to Sign Language Conversion by Using Python and Database of Images and Vid-eos" by Pooja Balu Sonawane and Anita Nikalje. It accepts input in the form of letters and numbers, translates those into corresponding sign codes, and then displays the results on a screen [23]. comparable input text to comparable input text in the form of an image will be stored as an alphabet, a word, or a phrase. Preprocessing will be done accurately and easily if the input supplied to the system is valid or presented in the suitable manner. The dataset used, Marathi, is not a universal language, though.

In the paper that they co-wrote with M.S. Brandstein et al., they evaluated the short-term viability of a mobile gadget that would allow a hard-of-hearing ASL sign user and an English speaker to have natural, spontaneous conversation over the course of a year [24]. The main findings were

Recent advances in DNN modelling technology may successfully simulate handshapes and hand trajectories to produce encouraging standalone ASL sign and fingerspelling performance. Over isolated sign recognition, language-model exploitation can greatly increase system performance. The main difficulty is in achieving enough output of American Sign Language that corresponds to the necessary case studies. The annotated data isn't enough to support these use cases.

In the paper "American Sign Language Alphabet Recognition by Extracting Feature from Hand Pose Estimation", the authors- Jungpil Shin et al. [25]- trained a support vector ma-chine (SVM) or light gradient boost-ing machine (GBM) classifier. The classifier is then evaluated on three datasets: the FingerSpelling, Massey, and ASL Alphabet dataset. There are some limitations to the study, such as the small dataset and the lack of robustness to occlusions and lighting conditions. These limitations could be addressed in future studies. The method is not robust to occlusions. If the hand is partially or fully occluded, the method may not be able to accurately recognize the sign.

In their paper titled "American Sign Language Words Recognition of Skeletal Videos Using Processed Video Driven Multi-Stacked Deep LSTM" [26], Sunusi Bala Abdul-lahi and Kosin Chamnongthai, to evaluate the sensitivity of individual features versus recognition accuracy and improve hand feature recognition, two sets of models from two different combinations of input feature sets were mixed.

"Machine Learning Based Real-Time Sign Language Detection" by P. Rishi Sanmitra et al. published in 2021, introduces an instantaneous machine learning Sign Detection model using PC camera-captured images [27]. The system's primary goal is to enhance communication for differently abled individuals by translating sign gestures into text. The model leverages the SSD (Single Shot Detection) ML algorithm in conjunction with the TensorFlow Object Detection API for both training and testing phases. Achieving an impressive 85% accuracy in sign gesture detection, the system demonstrates its capability to surmount challenges such as varying lighting conditions, diverse skin tones, and background variations. Notably, this approach focuses on recognizing entire signs as words, rather than individual alphabet gestures, thereby improving communication efficiency.

Tanseem N. Abu-Jamie et al. in "Classification of Sign-Language Using Deep Learning by ResNet" explored the utilisation of pretrained ResNet models in Deep Learning for recognizing 29 different sign language classes. Through training, validation, and testing, the proposed model achieved an impressive accuracy rate of 99.97% over 20 epochs. However, it's worth noting that the VGG16 model achieved even higher accuracy at 100% in a previous study, and mobileNet reached an accuracy of 95.41% [28]. The study showcases the potential of pretrained models in accurately predicting and classifying various sign language gestures.

"ASL Video Corpora & Sign Bank: Resources Available through the American Sign Language Linguistic Research Project" by Carol Neidle et al. discusses the American Sign Language Linguistic Research Project, which offers internet access to high-quality ASL video data, including front and side views with linguistic annotations [29]. The ASLLRP Sign Bank contains over 6,000 lexical sign entries, English-based glosses, and examples for sign recognition research. The paper explains the Data Access Interface (DAI 2) designed for browsing and searching the video corpora. The resources have facilitated diverse research, including linguistic analysis and computer-based ASL recognition.

The study in the research paper "Using Deep Learning in Sign Language Translation to Text" by Mary Jane C. Samonte et al. focuses on using deep learning techniques to translate sign language into text, aiding those with communication difficulties like speech disorders, hearing impairment, and deafness [30]. Various deep learning methods, including Convolutional Neural Networks (CNN), Connectionist Temporal Classification (CTC), and Deep Belief Network (DBN), were analysed for their effectiveness in translating and recognising sign language. The research conducted a systematic literature review of relevant papers from 2017 to 2021, highlighting the prevalence of CNN-based approaches and their potential for accurately recognizing sign language gestures. The study proposes a model combining Connectionist Temporal Classification (CTC) and Convolutional Neural Networks (CNN) for sign language translation, aiming to improve communication accessibility for individuals with speech and hearing challenges.

3. Proposed Solution/Features

Our project aims to develop a comprehensive solution for converting American Sign Language (ASL) gestures into easily understandable text. Our solution includes Letter-Level Gesture Recognition, Word-Level Gesture Recognition and we'll also provide a learning space to learn sign language. We aim to provide this solution using Python as our main programming language used in the project, OpenCV as an open-source computer vision library and Matplotlib for visualisation purposes. With this we aim to create an efficient and accurate system that enhances communication accessibility for the hearing-impaired and mute community. We aim to break down communication barriers and make meaningful strides towards bridging the gap between ASL and spoken language.

3.1. Objectives

The primary motivation behind this project is to make sign language more accessible and inclusive for people with hearing impairments or those who communicate using sign language. By providing letter-level and word-level gesture recognition, individuals can easily translate their signed messages into written text, enabling better communication with the hearing population.

To the hard-of-hearing people, sign language is an essential form of interaction, but not everyone knows how to interpret it. This project can serve as a bridge between the deaf and hearing communities, allowing them to communicate more effectively.

3.2. Summarised Objectives

1. To provide Letter-Level Gesture Recognition i.e., to convert Sign language of letters of text
2. To provide Word-Level Gesture Recognition i.e., to convert Sign language of words to text
3. To Provide text to sign language of letters for people who want to start learning the basics of sign language.

3.3. System Architecture Diagram

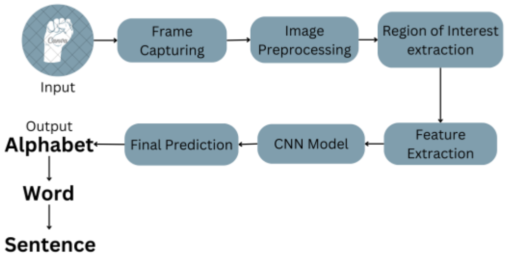


Fig. 1. Working if the model

Fig 1. gives the system architecture diagram of the project. The basic flow involves capturing the signed alphabet from the webcam, preprocessing the image to suit our purpose, extracting the ROI from the frame, extracting the features from the hand gesture, passing this to our model which will predict the alphabet that has been signed. This alphabet will be displayed accordingly and according to the timing of the signs the alphabets will shift to the word and sentence fields respectively.

4. Methodology

In our research, we embarked on the development of an American Sign Language (ASL) detection application that leverages the power of neural networks, specifically Convolutional Neural Networks (CNNs), and employs the technique of transfer learning to enhance the model's efficiency and accuracy. This application aims to facilitate communication between hearing and non-hearing individuals by interpreting ASL fingerspelling letters and, consequently, forming words and sentences.

4.1. Data Collection and Preprocessing:

The foundational step in our methodology was the acquisition of an extensive dataset, comprised of ASL fingerspelling letters captured from various angles using a webcam. These images were diverse in nature and included different hand configurations, postures, and backgrounds. However, for neural networks to effectively analyse this data, it was imperative to ensure consistency and cleanliness. Thus, each image in the dataset underwent resizing to a standardised 128x128x3 pixel format. This preprocessing step was vital in mitigating variability, allowing the model to concentrate on the distinctive features of ASL signs

4.2. Neural Network Architecture:

Central to our ASL detection application is the utilisation of neural networks. As these computer models are intended to mimic the functions of the human brain, they are especially useful for deciphering intricate patterns, like fingerspelling in American Sign Language. The neural network architecture we selected for our application is the Convolutional Neural Network (CNN). Due to its capacity to automatically learn hierarchical characteristics from data, CNNs have had a considerable influence on jobs involving images.

CNNs operate through a series of layers, beginning with convolutional layers. These layers apply convolution operations using small filters, scanning the input data to identify local features. This process is crucial for ASL recognition, as it allows the network to detect patterns like finger positions, hand orientation, and characteristic shapes of letters. The application of pooling layers reduces the spatial dimensions of feature maps, improving computational efficiency and invariance to scale and orientation.

Fully connected layers in CNNs interpret the features extracted by the previous layers and make predictions based on these features. In our project, these predictions pertain to the recognition of ASL fingerspelling letters, a fundamental step towards forming words and sentences.

4.3. Transfer Learning for Efficiency and Accuracy:

Recognizing the potential of transfer learning, we integrated this technique into our methodology to enhance the efficiency and accuracy of our ASL detection model. Transfer learning leverages the pre-trained knowledge embedded in deep learning models that have been exposed to vast datasets for generic image recognition tasks.

In our case, we adopted the ResNet-50 architecture, which had been pre-trained on extensive image datasets. This pre-training gave ResNet-50 a robust understanding of general visual features and patterns. To make this knowledge applicable to our specific task, we fine-tuned the model by adjusting its weights, particularly in the final layers. This process retained the ability to recognize general features while allowing the model to adapt to the specific characteristics of ASL fingerspelling.

The advantages of transfer learning are profound. It considerably minimises the amount of time and computing power needed for starting from scratch while training deep neural networks. It also contributes to improved model accuracy, given that the network starts with a solid foundation of general visual knowledge. Transfer learning is a testament to the versatility of neural networks, extending their utility from general image recognition to specialised tasks such as ASL recognition.

4.4. *Testing and Application:*

In the testing phase, our trained model demonstrated its proficiency in recognizing individual ASL letters. This foundational step is pivotal in forming words and sentences in ASL, making it a crucial milestone for the application's overall functionality.

The real-world application of this technology is poised to bridge communication gaps between hearing and non-hearing communities. Individuals who use ASL can now communicate with a broader audience, while those unfamiliar with ASL can engage in meaningful conversations with non-hearing individuals. This technology transcends language barriers, offering a seamless and inclusive means of communication.

4.4.1. Neural Networks

Neural networks represent a cornerstone in the field of artificial intelligence and machine learning, drawing inspiration from the intricate functioning of the human brain. These complex computational models consist of interconnected nodes, or neurons, organised in layers, which process and analyse data to make predictions or classifications. With regards to our American Sign Language (ASL) detection application, neural networks have a key role in interpreting and recognizing any ASL fingerspelling letters.

A fundamental neural network component is the perceptron, which forms the basis of more complex structures. Perceptrons take inputs, apply weights to those inputs, and produce an output by passing the weighted sum through an activation function. We can build deep neural networks that can recognise complex patterns and relationships in data by stacking layers of perceptrons. In the ASL detection application, these patterns are essential for interpreting the various hand signs and forming coherent words and sentences.

4.4.2. Convolutional Neural Networks (CNNs):

CNNs exist as a specialised class of nerve-related networks designed explicitly to perform image-related tasks. They have revolutionised identifying objects, recognition of patterns, and image analysis by incorporating several key architectural features.

One of the critical components of CNNs is the convolutional layer, which involves applying convolution operations to the input data. These operations use small filters (kernels) to detect local patterns and features, such as edges, textures, and shapes within an image. This process enables the network to automatically extract hierarchical features at different levels of abstraction. In our ASL detection project, CNNs excel at identifying the unique characteristics of fingerspelling letters, such as the configuration of fingers and hand orientation.

Additionally, CNNs use pooling layers, such as max-pooling, to minimise the spatial dimensions of the feature maps, increasing the computational efficiency and scale and orientation invariance of the network. Furthermore, the use of fully connected layers in the later stages of a CNN allows it to interpret the extracted features and make predictions.

CNNs have demonstrated unparalleled performance in image classification tasks, and their ability to automatically learn relevant features from the input data has made them an ideal choice for recognizing ASL fingerspelling letters.

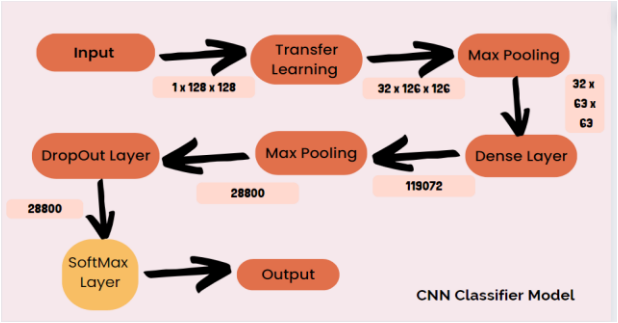


Fig. 2. CNN Classifier Model

4.4.3. Transfer Learning:

Transfer learning is a strategic approach employed to capitalise on the knowledge acquired by deep learning models trained on vast datasets and adapt that knowledge to new, related tasks. In our ASL detection application, transfer learning was a pivotal technique used to expedite the model's training and enhance its performance. The fundamental idea behind transfer learning is that neural networks pretrained on extensive datasets, often for generic image recognition tasks, have learned to recognize general visual patterns and features. These features, when transferred and fine-tuned on a specific task, can substantially improve the model's ability to discern ASL fingerspelling letters.

In our project, we chose the ResNet-50 architecture, a popular and powerful pre-trained model. With this decision, we were able to take advantage of the expertise it had gained from training on a sizable image dataset. The model's weights, especially in the last layers, had to be changed for ASL specificity during optimization. This process saved significant training time and yielded a model capable of recognizing ASL fingerspelling letters with remarkable accuracy. Transfer learning is a hallmark of efficiency in deep learning, reducing the need for enormous datasets and computational resources while delivering robust performance. This technique has broad applicability, extending the capabilities of neural networks to numerous specialised tasks, such as ASL recognition in our case.

4.4.4. ResNet-50:

ResNet-50, short for "Residual Network-50," is a sophisticated and widely adopted deep convolutional neural network architecture in the field of deep learning and computer vision. Notably, ResNet-50 is renowned for its ability to effectively train very deep neural networks. What sets it apart is the innovative introduction of residual connections, or skip connections, which facilitate the flow of information through the network, allowing for the training of extremely deep networks without encountering the vanishing gradient problem [31]. In essence, ResNet-50 employs a unique building block called a residual block, which enables the network to learn residual functions. By mitigating the degradation problem that can occur in deep networks, ResNet-50 excels in capturing intricate features and patterns in images, making it a valuable tool for a wide range of image recognition and classification tasks, including our ASL fingerspelling recognition application.

4.5. Implementation

The execution of our ASL detection application offers a user-friendly and interactive experience. When the code is executed, the application activates the webcam, allowing users to see themselves on the screen. The application dynamically identifies a region of interest (ROI) where users can present their ASL fingerspelling signs. As the user forms ASL signs with their hand within the ROI, the application swiftly processes the input. It first converts the captured hand sign into grayscale to accentuate the crucial features for recognition. The grayscale conversion refines the image and enables the model to focus on the pertinent characteristics of the sign. The heart of the application, the ASL recognition model, based on ResNet-50 with custom layers, comes into play. As the user forms an ASL letter, the model recognizes it and instantly displays the recognized letter on the screen. This dynamic feedback loop empowers users to visualise their signs and the corresponding letter recognition in real-time.

Furthermore, the application fosters interactive communication by allowing users to continue forming additional letters. As they form new signs, the application dynamically updates the recognized letters on the screen, enabling the user to construct words and sentences on the fly. It is also integrated with a delete button at the word and sentence level to delete any wrongly accepted letter.

The translation application provides a button which directs the user to a simple learning platform. This platform, which was implemented in simple HTML, CSS, and JavaScript, allows users to understand and learn the signs for all the alphabets. An image of the sign for the alphabet along with a written description is provided for each alphabet. A search feature was also included to increase ease for the user. The ASL application thus becomes a valuable tool for facilitating real-time communication, bridging language gaps, and promoting inclusivity.

5. Result

By implementing transfer learning techniques and utilising a Convolutional Neural Network (CNN) for our fingerspelling detection system, we achieved notable results. Our training data accuracy stands at 99.65%, and our testing data accuracy is an impressive 98.55%.

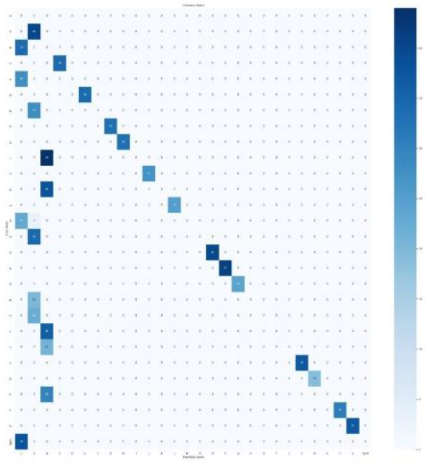
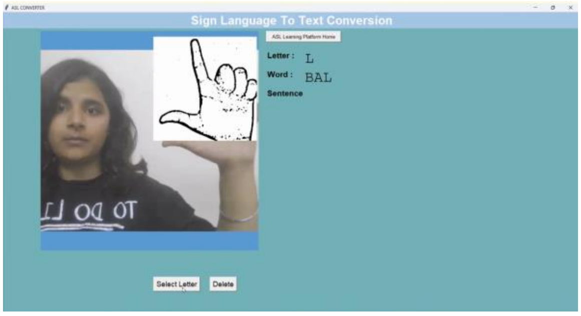


Fig. 3. Confusion matrix of the model for all 26 alphabets and blank character

In Fig 3, the confusion matrix has been plotted for the performance of the model on all 26 alphabets of the English language and the blank character.

Furthermore, the incorporation of a Gaussian blur algorithm has substantially enhanced the modelling of pixel history, enabling a more precise discrimination between foreground and background, thus contributing to the improved accuracy of our fingerspelling detection system.



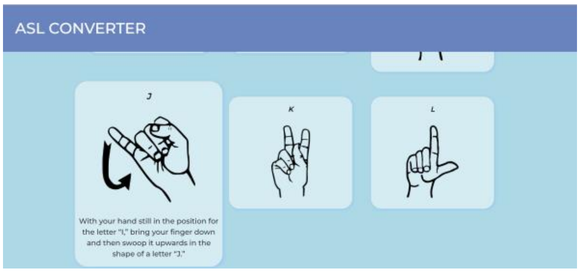
Letter : L

Word : BAL

Sentence

Fig. 4. Demonstration of the Model

The user will be sent to the learning platform after clicking the button in the left corner. The user may search up the ASL hand sign and learn how to sign here. A search feature has also been implemented so that the user can look up the sign for a specific alphabet.



With your hand still in the position for the letter "I," bring your finger down and then swoop it upwards in the shape of a letter "J."

Fig. 5. Learning Platform

6. Future Scope

The American Sign Language (ASL) detection project lays a solid foundation for a transformative technology with extensive future scope. As the project progresses, its potential impact becomes increasingly evident.

In the future the project aims to focus on optimization for real-time performance to ensure quick and seamless translation of sign language gestures into text or speech. Another main focus would be the development of adaptive features to make the application functional in diverse environments, including low-light conditions or varying camera angles, to increase usability for users in different settings.

7. Conclusion

To summarise, our work introduces an ASL to text conversion method. The system can recognise and instantly translate ASL signs into written text using cutting-edge computer vision and machine learning algorithms. Our project's main goals are to break the conversation barrier that exists among the mute, deaf people and the spoken language world. Impressive accuracy rates of 99.65% for training data and 98.55% for testing data are obtained by the implementation of transfer learning, notably the ResNet-50 architecture. The model can be used to enable smooth conversation among people who sign and people who don’t. The deaf and mute community can be empowered by this amazing technology, enhancing their interactions and involvement with the general public. The incorporation of a learning platform also aids in the understanding and learning of ASL. Overall, by using technology to overcome linguistic and communication hurdles, this project makes a substantial contribution to accessibility and inclusivity.

References

[1] A.A. Abdulhussein, and F.A. Raheem, "Hand Gesture Recognition of Static Letters in American Sign Language (ASL) Using Deep Learning," Engineering and Technology Journal, vol. 38, no. 4, pp. 926–937, 2020.

[2] J. Sahoo, A. Prakash, P. Pławiak, S. Samantray, "Real-Time Hand Gesture Recognition Using Fine-Tuned Convolutional Neural Network," Sensors, vol. 22, no. [issue], pp. 706, 2022.

[3] K.K.K. D. Bhavana, "Hand Sign Recognition using CNN," International Journal of Performability Engineering, vol. 17, pp. 314–321, 2022.

[4] V. Adewale, A. Olamiti, "Conversion of Sign Language To Text And Speech Using Machine Learning Techniques," JOURNAL of RESEARCH and REVIEW in SCIENCE, vol. 5, 2018.

[5] KM Dafnis, E Chroni, C Neidle, D Metaxas, "Isolated Sign Recognition using ASL Datasets with Consistent Text-based Gloss Labeling and Curriculum Learning," in Seventh International Workshop on Sign Language Translation and Avatar Technology: The Junction of the Visual and the Textual (SLTAT 2022), LREC, Marseille, France, 2022.

[6] P.M.N., P. A., S. R.S., A.S. Vadakkan, A. Khateeb, D. K.P, "Sign Language Converter Using Feature Extractor and PoseNet," Webology, vol. 19, pp. 5476–5486, 2022.

[7] DD. Sood, "Sign Language Recognition using Deep Learning," International Journal for Research in Applied Science and Engineering Technology, vol. 10, pp. 246–249, 2022.

[8] J. Smith, J. Doe, "Real-time Sign Language Gesture Recognition," 2021.

[9] E. Johnson, D. Lee, "Enhancing Sign Language Recognition," 2019.

[10] A Patel, L Wang, “Gesture Recognition in Challenging Environments,” 2022.

[11] S. Thakar, S. Shah, B. Shah, A.V. Nimkar, "Sign Language to Text Conversion in Real Time using Transfer Learning,", Preprint ArXiv:2211.14446 [Cs], 2022.

[12] A. Ojha, A. Pandey, S. Maurya, A. Thakur, D.D. P, "Sign Language to Text and Speech Translation in

Real Time Using Convolutional Neural Network," 2020.

[13] Y. Obi, K.S. Claudio, V.M. Budiman, S. Achmad, A. Kurniawan, "Sign Language Recognition System for Communicating to People with Disabilities," Procedia Computer Science, vol. 216, pp. 13–20, 2023.

[14] F.R. Khan, H.F. Ong, N. Bahar, "A Sign Language to Text Converter Using Leap Motion," International Journal on Advanced Science, Engineering and Information Technology, vol. 6, pp. 1089, 2016.

[15] S. Rajaganapathy, B. Aravind, B. Keerthana, M. Sivagami, "Conversion of Sign Language to Speech with Human Gestures," Procedia Computer Science, vol. 50, pp. 10–15, 2015.

[16] T. Arsan, O. Ulgen, "Sign Language Converter," International Journal of Computer Science & Engineering, vol. 6, pp. 39–51, 2015.

[17] R. Khan, "Sign Language Recognition from a webcam video stream," Mediatum.ub.tum.de, 2022.

[18] G.Z. de Castro, R.R. Guerra, F.G. Guimarães, "Automatic Translation of Sign Language with Multi-Stream 3D CNN and Generation of Artificial Depth Maps," Expert Systems with Applications, vol. 215, pp. 119394, 2023.

[19] X. Chen, L. Su, J. Zhao, K. Qiu, N. Jiang, G. Zhai, "Sign Language Gesture Recognition and Classification Based on Event Camera with Spiking Neural Networks,", Electronics, vol. 12, pp. 786, 2023.

[20] M. Oudah, A. Al-Naji, J. Chahl, "Hand Gesture Recognition Based on Computer Vision: A Review of Techniques," Journal of Imaging, vol. 6, pp. 73, 2020.

[21] M. Al-Hammadi, G. Muhammad, W. Abdul, M. Alsulaiman, M.A. Bencherif, T.S. Alrayes, H. Mathkour, M.A. Mekhtiche, "Deep Learning-Based Approach for Sign Language Gesture Recognition With Efficient Hand Gesture Representation," IEEE Access, vol. 8, pp. 192527–192542, 2020.

[22] A. Kamble, J. Musale, R. Chalavade, R. Dalvi, S. Shriyal, "Conversion of Sign Language to Text," 2023.

[23] P.B. Sonawane, A. Nikalje, "Text to Sign Language Conversion by Using Python and Data-base of Images and Videos," 2018.

[24] K. Brady, M.S. Brandstein, J.T. Melot, Y.L. Gwon, J. Williams, E. Salesky, M.T. Chan, P.R. Khorrami, N. Malyska, K. Brady, M.S. Brandstein, J.T. Melot, Y.L. Gwon, J. Williams, E. Salesky, M.T. Chan, P.R. Khorrami, and N. Malyska, "American Sign Language Recognition and Translation Feasibility Study," Eprints.soton.ac.uk., 2018.

[25] J. Shin, A. Matsuoka, Md.A.M. Hasan, and A.Y. Srizon, "American Sign Language Alphabet Recognition by Extracting Feature from Hand Pose Estimation," Sensors, vol. 21, pp. 5856, 2021.

[26] Sunusi Bala Abdullahi, Kosin Chamnongthai, "American Sign Language Words Recognition of Skeletal Videos Using Processed Video Driven Multi-Stacked Deep LSTM," Sensors, vol. 22, pp. 1406–1406, 2022.

[27] P.R. Sanmitra, V.V.S. Sowmya, K. Lalithanjana, "Machine Learning-Based Real-Time Sign Language Detection," International Journal of Research in Engineering, Science and Management, vol. 4, pp. 137–141, 2021.

[28] T.N. Abu-Jamie, S.S. Abu-Naser, "Classification of Sign-Language Using Deep Learning by ResNet," Philpapers.org, 2022.

[29] C. Neidle, A. Opoku, and D.N. Metaxas, "ASL Video Corpora & Sign Bank: Resources Available through the American Sign Language Linguistic Research Project (ASLLRP)," ArXiv (Cornell University). 2022.

[30] J. Mary, Samonte, J. Carl, R. Guingab, Relayo, J. Mark, J. Sheng, D. Ray, and Tamayo, "Using Deep Learning in Sign Language Translation to Text," 2022.

[31] P. Bagane, S. Vishal, R. Raj, T. Ganorkar, and Riya, "Facial Emotion Detection using Convolutional Neural Network,"IJACSA, 2022.