**Abstract**:

**Introduction**:

The human language is a powerful device that allows us to communicate our ideas, share our thoughts and feelings, and express our viewpoints. What we take for granted plays a critical role in almost all human interactions. Most of our daily activities assumes that a person is able to use the spoken human language. Therefore, it is easy for us to overlook the struggle of people with hearing disabilities such as the mute, deaf or hard of hearing who have trouble understanding and expressing themselves to people who are different from them. Sign language is supposed to act as the bridge to connect people with hearing disabilities to connect with others. Contrary to popular to belief, there is no universal sign language as each sign language is distinct to a particular human language. For this project we will focus on ASL, a complete natural language that has the same linguistic properties as spoken languages and is primarily expressed through the use of hand gestures. There is a limited amount of people who are proficient in american sign language so there is a big gap between the number of interpreters and the population with hearing impairment. Although the use of interpreters can help facilitate communication, it often makes them overly reliant or dependent on the interpreter. In this project, we aimed to create a american sign language recognition system that can be used as a tool to bridge the gap between hearing-impaired community and those untrained in American Sign Language (ASL). Ultimately, we hope our project can empower hearing impaired population to be more independent and improve the comfort in everyday human interactions.

**Related Works**:

We are not the first to recognize the potential benefits of a sign language recognition system to the population of people with hearing disabilities. Previous literature have tackled the problem of sign recognition in dynamic videos and static images using a variety of machine learning methods and algorithms.

A popular approach to solve this problem is to use a HMM model, a probabilistic model that is based on the statistical Markov model to recognize the relationship between observable events that depend on internal factors. Previous literature has shown success with using an HMM model to recoginze sign languages. Thad Starner and Alex Pentland created a real time system that interpreted American Sign Language using Hidden Markov Models. The system tracked a person’s hand by their shape, orientation, and gesture which served as the input to the HMM model to recognize the signed gesture. The research group employed two experiments. The first experiment tracked hands wearing colored gloves while the second experiment used hands without gloves. Ultimately, they found their system achieved a higher accuracy when tracking hands with gloves [1].

With the rising popularity of neural networks and deep learning, more recent work explore how deep learning approaches compare to those of the past. The most traditional approach to solve computer vision related problems is CNN that utilizes the features of an image to aid in classification. Machine learning models with CNN like architectures have shown some success in this topic.

Simming He proposed a deep learning approach which consisted of using a combination of methods. He used Faster R-CNN with embedded RPN module to locate the hands in the images and feeds the results into a 3D CNN feature extraction network and sign language recognition framework based on long and short time memory(LSTM) coding and decoding network[2].

Rabeet Fatmi et al. created a system that utilized wearable motion sensors and Aritifical Neural Networks and Support Vector Machines models to recognize words in ASL. The research showed high accuracy in ANN compared other machine learning approaches. [3]

Lionel Pigou et al. built a real time system to recognize 20 italian gestures using Microsoft Kinect, CNN, and GPU acceleration. The system utilized CNN model to extract features from the frames and used an artificial neural network for the classification [4].

**Methodology**:

Solving machine learning problems can be categorized into the following steps: task definition, data collection, exploratory data analysis, machine learning modeling, model testing, and deployment.

**Problem Definition**: We want to create a ASL recognition system using supervised machine learning algorithms. More specifically, given a image of an american sign language gesture, we want to system to be able to recognize the sign.

**Data Collection**: To solve any machine learning related task, it is important to collect a dataset that is representative of the problem you are trying to solve. Due to the limited time of the assignment and the lack of expertise in ASL, we could not create our own comprehensive dataset. Instead we had to make the best out of the publicly available datastets for this problem. We used the popular Sign Language MNIST dataset that consisted of 27,455 training and 7,172 testing grayscale images of size 28 x 28. Each labeled images represented an american alphabet letters besides J and Z that required more dynamic gestures that can not be captured in static images.

**Exploratory Data Analysis**: EDA is an important step when creating machine learning models because it provides insight into the data and allows you recognize the relevant and irrelevant features. We learned that our data is balanced and extremely clean because the dataset was only composed of cropped hands.

**Machine Learning Modeling:** To create a ASL recognition system we employed various supervised machine learning techniques such as CNN, ResNet50, InceptionV3, and YoloV3 to find the one that works best for this problem.

Convulational neural network is a classical approach used to solve computer vision related problems. CNN is a neural network model that consists of multiple convolutional layers, flatten layers, dense layers, and finally a softmax activation function that provides the classification. CNN is a powerful machine learning algorithm because it can automatically detect and learn the important features to classify an image without any human supervision. Additionally, they can detect features anywhere in the image whereas it’s predecessors had difficulties learning features that were in spatially different locations [5].

// Attach image of CNN Architecture to support the blurb above ^

Resnet50 here

InceptionV3 here

YoloV3 here

**Results**:

**Discussion**:

* Discuss limitations of sign language that a person loses information on the tone and other vocal indicators that provide meaning.

**\*\*Future Works**:

* Text to sound
* Word level recognition

**Acknowledgements**:

**References**:

1. <https://www.aaai.org/Papers/Symposia/Fall/1996/FS-96-05/FS96-05-017.pdf>
2. <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8950864&casa_token=Giaeh8_NHkcAAAAA:yVMdbRTf_z59E73fJd74H2a4NmVSD_CuFTTrXkBxc8ZE91k691TMp48CQygGJ063-kxnqxVb&tag=1>
3. <http://www.ibai-publishing.org/journal/issue_mldm/2017_October/10_2_41_55_mldm.pdf>
4. <https://link.springer.com/content/pdf/10.1007/978-3-319-16178-5_40.pdf>
5. <https://insightsimaging.springeropen.com/articles/10.1007/s13244-018-0639-9>
6. <https://www.nidcd.nih.gov/health/american-sign-language>
7. <https://www.ijert.org/sign-language-recognition-system-using-convolutional-neural-network-and-computer-vision#:~:text=We%20make%20use%20of%20Convolutional,remarkable%20accuracy%20of%20above%2090%25.&text=recognition%2C%20Sign%20language%20recognition%2C%20Hue%20Saturation%20Value%20algorithm>.