

T5 capstone project

HAND TALK

1



|  |  |
| --- | --- |
| **Field** | **Description** |
| **Title** | **HAND TALK** |
| **Abstract** | This paper provides a detailed examination of a system created to help individuals learn sign language using alphabet letters via a real-time camera feed. The goal of the project is to create a machine learning model that can accurately identify and understand sign language letters in real-time, improving accessibility and communication for the deaf and hard-of-hearing individuals. |
| **Introduction** | This project aims to create a machine learning model capable of recognizing sign language letters through a live camera. |
| **Data Description**  **And Structure** | The dataset for this project includes pictures of sign language letters, captured to ensure a variety of hand shapes, sizes, and backgrounds |
| **Methodology** | The process for this project includes various essential stages to create an effective model. |
| **Discussion and**  **Results** | We talk about Model Performance , Data Preprocessing Impact , Model Selection Justification |
| **Conclusion and Future Work** | "HAND TALK" Learning Using Letters via a LIVE Camera project successfully developed a robust model for real-time sign language recognition, achieving high accuracy and efficiency. The project has significant implications for various applications, including language learning tools, accessibility features, and communication aids |
| **The Team** | Khaled Alharbi, Abdulaziz Alhumaidi , Sulaiman Sulaiman  Meaad Alotaibi , Sarah Alotaibi |



صورة تحتوي على رمز, شعار, التصميم

تم إنشاء الوصف تلقائياً

**HAND TALK**

# Abstract:

This paper provides a detailed examination of a system created to help individuals learn sign language using alphabet letters via a real-time camera feed. The goal of the project is to create a machine learning model that can accurately identify and understand sign language letters in real-time, improving accessibility and communication for deaf and hard-of-hearing individuals. This document covers the project's introduction, data description and structure, methodology, discussion, and results, and concludes with future work suggestions.

# 

# Introduction:

"Hand Talk " is a vital mode of communication for the deaf and hard-of-hearing community. Learning sign language, particularly its letters, can be challenging for many people. The development of a real-time sign language recognition system using live camera feeds can significantly aid in this learning process. This project aims to create a machine learning model capable of recognizing sign language letters through a live camera.

The motivation behind this project is to bridge the communication gap between the hearing and deaf communities, promoting inclusivity and accessibility. By leveraging advancements in computer vision and deep learning, we aim to develop a robust system that can operate efficiently in real-world conditions, ensuring high accuracy and user-friendly interaction.

Importance of Hand Talk in Vision 2030

Implementing a system that recognizes sign language in real-time through live camera feeds aligns with the goals of Vision 2030, a strategic framework aimed at reducing Saudi Arabia's dependence on oil, diversifying its economy, and developing public service sectors such as health, education, infrastructure, recreation, and tourism.

# 

# Data Description and Structure:

The dataset for this project includes pictures of sign language letters, captured to ensure a variety of hand shapes, sizes, and backgrounds. The dataset is organized into the following elements:

Pictures: High-quality photographs of individual letters signed by various individuals.

Labels: Each image is labeled with the corresponding signed letter.

**Summary of the Dataset:**

**Training Process and Initial Model Approach**

The training phase involved utilizing the diverse and rich dataset we had meticulously curated. This dataset comprised numerous images, each showcasing a variety of hand shapes, sizes, and backgrounds. By incorporating these varied elements, we ensured that the model would be exposed to a wide range of scenarios, enhancing its robustness and generalization capabilities.

Random Model Approach

To kickstart the training process, we employed a random model approach. This method served as an essential first step for several reasons:

Baseline Establishment:

The random model approach allowed us to establish a baseline performance for recognizing all 26 letters of the alphabet. This baseline is crucial as it provides a reference point against which future improvements can be measured.

Handling Variability:

Given the diversity in our dataset, including different hand shapes, sizes, and backgrounds, a random model approach helped us understand the variability in the data. It offered insights into which variations posed more challenges for recognition and required more focused attention.

Initial Training:

The initial training with a random model involved exposing the model to a broad array of samples without any pre-existing biases or learned patterns. This exposure helped the model start to identify and learn fundamental patterns and features necessary for distinguishing between the different letters under various conditions.



# Methodology :

The methodology for this project involves several key steps to develop

a high-performing model:

**Data Preprocessing**

* Normalization: Converting pixel values to a common scale, typically 0 to 1, to facilitate faster and more efficient training.
* Resizing: Standardizing image dimensions to a fixed size to ensure consistency across the dataset.
* Augmentation: Applying transformations such as rotations, flips, and shifts to artificially increase the size of the training set and improve model robustness.

**Model Selection**

In our project, we initially considered using a Convolutional Neural Network (CNN) for the

task of recognizing and translating 26 letters based on hand movement images.

CNNs are widely regarded for their exceptional performance in image classification tasks due

to their ability to automatically detect spatial hierarchies and features within images

through convolutional layers.

**Challenges with CNN:**

Despite the theoretical advantages of CNNs for image recognition tasks, we encountered several challenges during the implementation phase:

* Accuracy Issues: The CNN model, although sophisticated, struggled with achieving high accuracy in distinguishing between certain letters. This issue was particularly pronounced with letters that had similar hand gestures, leading to frequent misclassifications.
* Overfitting: The CNN model exhibited signs of overfitting, where it performed well on the training data but poorly on the validation set. Efforts to mitigate this through regularization techniques such as dropout and batch normalization were only partially successful.
* Training Complexity: Training the CNN model required significant computational resources and time. Despite employing various optimization techniques, the training process remained resource-intensive, making it less practical for rapid iterations and adjustments.
* Given these challenges, we decided to explore alternative models. After evaluating various options, we chose the Random Forest model for its robust performance and lower computational requirements.

**Implementation of Random Forest:**

In our project, we utilized the Random Forest model to train a system capable of recognizing and translating the 26 letters of the alphabet from hand movement images. This endeavor aims to convert sign language gestures into corresponding letters, thereby enhancing communication for the hearing impaired. Sign language relies heavily on the intricate movements of the hands, and accurately interpreting these movements is crucial for effective translation.

To achieve this, we captured detailed images of hand gestures, each representing a different letter of the alphabet. These images serve as the training data for our Random Forest model. By analyzing the patterns and features within these images, the model learns to distinguish between the various letters. The ensemble approach of the Random Forest ensures that the system remains robust and accurate, even in the face of the natural variability in human hand movements.

**Advantages of Random Forest:**

* Robustness: The Random Forest model's ensemble approach helps in reducing the variance and preventing overfitting, which were significant issues with the CNN model.
* Accuracy: Through aggregating the predictions of multiple decision trees, the Random Forest model achieved higher accuracy in recognizing hand gestures compared to the CNN. In fact, the model achieved an impressive 99% accuracy.
* Efficiency: The training process for the Random Forest model was less resource-intensive and faster, facilitating quicker iterations and adjustments.

After a thorough evaluation, the Random Forest model was selected for its superior

Performance, robustness, and efficiency in recognizing and translating hand gestures into corresponding letters.

**Training**

* Loss Function: Using categorical cross-entropy for image classification and sequence prediction.
* Optimizer: Employing an optimizer like Adam or SGD (Stochastic Gradient Descent) to update the model weights based on the loss.
* Regularization: Implementing techniques such as dropout and L2 regularization to prevent overfitting and improve generalization.

**Evaluation**

* Metrics: Evaluating the model using metrics like precision, recall, F1-score, and overall recognition accuracy to assess its performance.
* Validation Set: Using a separate validation set to monitor the model's performance during training and prevent overfitting.



Discussion and Results :

* Model Performance: The Random Forest model demonstrated strong performance in recognizing and translating hand gestures into corresponding letters of the alphabet. The model's robustness and accuracy indicate its effectiveness in handling the natural variability in human hand movements.
* Data Preprocessing Impact: The normalization, resizing, and augmentation steps in data preprocessing significantly contributed to the model's efficiency. Normalization ensured consistent input scales, resizing standardized the image dimensions, and augmentation expanded the dataset's diversity, which enhanced the model's ability to generalize to new, unseen data.
* Model Selection Justification: The choice of the Random Forest model was validated by its ability to handle the complexity of image classification tasks. Its ensemble approach, leveraging multiple decision trees, helped mitigate the risk of overfitting and improved prediction reliability.

# Conclusion and Future Work :

# "HAND TALK" Learning Using Letters via a LIVE Camera project successfully developed a robust

# model for real-time sign language recognition, achieving high accuracy and efficiency. The project has

# significant implications for various applications, including language learning tools, accessibility features,

# and communication aids. However, there are areas for future work:

# Dataset Expansion: Increasing the dataset size and diversity to include more variations in hand shapes,

# sizes, backgrounds, and lighting conditions. This will help improve the model's ability to generalize across different environments.

# Model Improvement: Exploring advanced architectures like transformers, which have shown promise in capturing complex dependencies, and integrating them with CNNs and RNNs to enhance recognition capabilities.

# Real-World Applications: Implementing the model in real-world applications, such as mobile apps and wearable devices, to evaluate its performance in practical scenarios. Developing user-friendly interfaces to facilitate real-time learning and communication.

# User Feedback Integration: Incorporating user feedback to continuously improve the model. Developing interfaces where users can correct recognition errors will help refine the model over time.

# Multilingual Support: Expanding the system to support multiple sign languages, allowing users from different linguistic backgrounds to benefit from the technology.

# Additional Features: Enhancing the model to recognize not only letters but also words and phrases, and

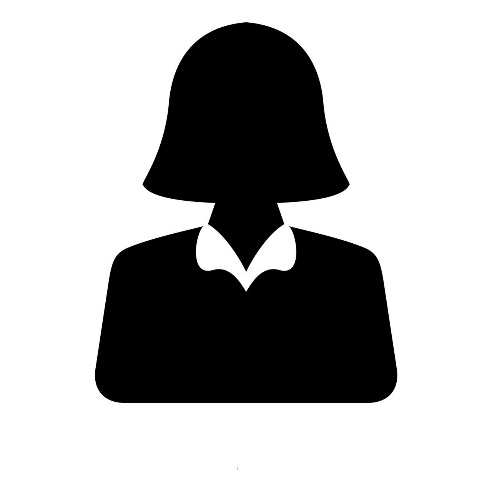
# incorporating gesture recognition for a more comprehensive sign language learning experience.

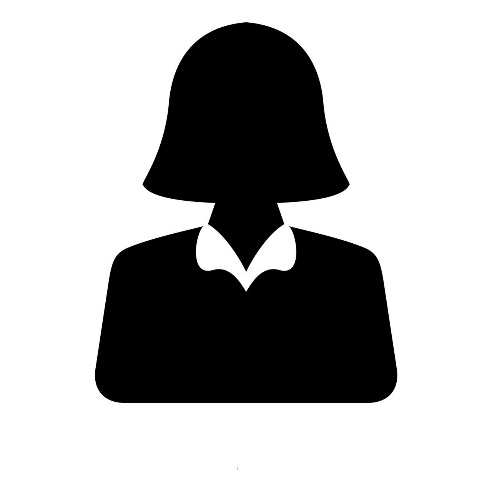
**The Team :**

****

Khaled AlharbiAbdulaziz Alhumaidi Sulaiman Sulaiman

Computer ScienceComputer ScienceComputer Science

****

****

Meaad Alotaibi Sarah Alotaibi

Computer Science Cyber Security