##### A Project report on

**GesturePath:Real-Time Sign Language Detection with Action Recognition**

###### A Dissertation submitted to JNTUH, Hyderabad in partial fulfillment of the academic requirements for the award of the degree.

**Bachelor of Technology**

**in**

**Artificial Intelligence and Machine Learning**

Submitted by

G. MANOJ

(21H51A7303)

M. DEVAVRATH

(21H51A7305)

R. SIRI

(21H51A7308)

Under the esteemed guidance of

Mr. Enoch Raja. DG(Assistant Professor)



**Department of Artificial Intelligence and Machine Learning**

**CMR COLLEGE OF ENGINEERING& TECHNOLOGY**

(UGC Autonomous)

\*Approved by AICTE \*Affiliated to JNTUH \*NAAC Accredited with A+ Grade

KANDLAKOYA, MEDCHAL ROAD, HYDERABAD - 501401.

#### 2024-2025

.

**CMR COLLEGE OF ENGINEERING & TECHNOLOGY**

KANDLAKOYA, MEDCHAL ROAD, HYDERABAD – 501401

**DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING**



#### CERTIFICATE

This is to certify that the Major Project Phase-1 report entitled **"GesturePath:Real-Time Sign Language Detection with Action Recognition"** being submitted by G.MANOJ (21H51A7303), M.DEVAVRATH (21H51A7305), R.SIRI (21H51A7308) in partial fulfillment for the award of **Bachelor of Technology in Artificial Intelligence and Machine Learning** is a record of bonafide work carried out his/her under my guidance andsupervision.

###### The results embodies in this project report have not been submitted to any other University or Institute for the award of any Degree.

**Mr. Enoch Raja. DG Dr. S. Kirubakaran**

**(Assistant Professor) Professor and HOD**

**Dept. of AIML Dept. of AIML**

#### ACKNOWLEDGEMENT

With great pleasure we want to take this opportunity to express my heartfelt gratitude to all the people who helped in making this project work a grand success.

We are grateful to **Mr. Enoch Raja. DG,** **Assistant Professor**, Department of Artificial Intelligence and Machine Learning for his valuable technical suggestions and guidance during the execution of this project work.

We would like to thank **Dr. S. Kirubakaran,** Head of the Department of Artificial Intelligence and Machine Learning, CMR College of Engineering and Technology, who is the major driving forces to complete my project work successfully.

We would like to thank **Dr. P. Ravi Kumar**, Dean F&S, CMR College of Engineering and Technology, for his insight and expertise have been instrumental in shaping the direction and execution of this project work successfully.

We are very grateful to **Dr. Ghanta Devadasu**, Dean-Academics, CMR College of Engineering and Technology, for his constant support and motivation in carrying out the project work successfully.

We are highly indebted to **Dr. V A Narayana,** Principal, CMR College of Engineering and Technology, for giving permission to carry out this project in a successful and fruitful way.

We would like to thank the Teaching & Non- teaching staff of Department of Computer Science and Engineering for their co-operation

We express our sincere thanks to **Shri. Ch. Gopal Reddy**, Secretary& Correspondent, CMR Group of Institutions, and **Shri Ch Abhinav Reddy**, CEO, CMR Group of Institutions for their continuous care and support

Finally, We extend thanks to our parents who stood behind us at different stages of this Project. We sincerely acknowledge and thank all those who gave support directly and indirectly in completion of this project work.

G. MANAOJ 21H51A7303

M. DEVAVRATH 21H51A7305

R. SIRI 21H51A7308

GesturePath:Real-Time Sign Language Detection with Action Recognition

**TABLE OF CONTENTS**

**CHAPTER**

**NO. TITLE PAGE NO.**

LIST OF FIGURES ii

LIST OF TABLES iii

ABSTRACT iv

**1** **INTRODUCTION** 1

1.1 Problem Statement 2

1.2 Research Objective 2

1.3 Project Scope and Limitations 3

**2** **BACKGROUND WORK** 4

2.1. Static Image-Based Methods 5

2.1.1.Introduction 5

2.1.2.Merits,Demerits and Challenges 5

2.1.3.Implementation of Static Image-Based Methods 6

2.2. Dynamic Gesture Recognition (HMMs) 7

2.2.1.Introduction 7

2.2.2.Merits,Demerits and Challenges 7

2.2.3.Implementation of Dynamic Gesture Recognition (HMMs) 8

2.3. Deep Learning-Based Recognition (CNN) 9

2.3.1.Introduction 9

2.3.2.Merits,Demerits and Challenges 9

2.3.3.Implementation of Deep Learning-Based Recognition (CNN) 10

**3 RESULTS AND DISCUSSION** 11 3.1. Comparison of Existing Solutions 12

3.2. Data Collection and Performance metrics 13

**4** **CONCLUSION** 14

6.1 Conclusion 15

**5** **REFERENCES** 17

CMRCET B.Tech (AIML) PageNo i

GesturePath:Real-Time Sign Language Detection with Action Recognition

**List of Figures**

**FIGURE**

**NO. TITLE PAGE NO.**

2.1 Static Image-Based Recognition 6

2.2 Dynamic Gesture Recognition8

2.3 Deep Learning-Based Sign Language Recognition 10

CMRCET B.Tech (AIML) PageNo ii

GesturePath:Real-Time Sign Language Detection with Action Recognition

**List of Tables**

**FIGURE**

**NO. TITLE PAGE NO.**

### 3.1 **Comparison of Existing Solutions** 11

#### 3.3 **Performance Metrics** 12

CMRCET B. Tech (AIML) PageNo iii

GesturePath:Real-Time Sign Language Detection with Action Recognition

# **ABSTRACT**

Communication barriers faced by the deaf and mute community often lead to social isolation. The project titled **"GesturePath: Real-Time Sign Language Detection with Action Recognition"** aims to bridge this gap by converting sign language gestures into text using a Long Short-Term Memory (LSTM) neural network. Utilizing MediaPipe Holistic for keypoint detection, the system preprocesses gesture data to extract meaningful patterns for action recognition. This project provides a user-friendly and efficient solution to facilitate seamless communication.

The system employs MediaPipe Holistic for detecting and extracting keypoints from hand, face, and body landmarks, capturing intricate gesture movements. This data is then processed using a Long Short-Term Memory (LSTM) neural network, which excels in learning temporal dependencies essential for action recognition.

The pipeline includes steps such as preprocessing raw data, training the model on labeled datasets, and real-time gesture prediction. Through rigorous evaluation, the system demonstrated high accuracy in recognizing dynamic gestures, underscoring its potential to enhance communication accessibility for the deaf and mute. This project contributes to the growing field of assistive technology by offering an inclusive solution that fosters understanding and bridges communication divides.

CMRCET B.Tech (AIML) PageNo iv

GesturePath:Real-Time Sign Language Detection with Action Recognition

# **CHAPTER 1**

**INTRODUCTION**

CMRCET B.Tech (AIML) PageNo 1

GesturePath:Real-Time Sign Language Detection with Action Recognition

**CHAPTER 1**

**INTRODUCTION**

### 1.1 Problem Statement

Communication is a vital aspect of human interaction, but the lack of understanding of sign language often limits accessibility for the deaf and mute. The GesturePath system focuses on converting sign language into text, making it easier for non-sign language users to communicate effectively.

This project leverages keypoint detection techniques using MediaPipe Holistic and processes them using a machine learning model. The steps include keypoint extraction, data preprocessing, and the development of an LSTM neural network for gesture recognition. With an accuracy-focused approach, this project aims to provide real-time gesture detection and recognition.

### 1.2 Research Objective

The primary objective of this research is to develop a robust and efficient real-time system capable of translating sign language gestures into textual representations. The key goals include:

1. Leveraging MediaPipe Holistic to detect and extract keypoints from hand, face, and body movements.
2. Building an LSTM neural network model to accurately interpret temporal dependencies in gesture sequences.
3. Ensuring high accuracy in gesture recognition through extensive training and testing on diverse datasets.
4. Designing a user-friendly interface that facilitates seamless communication between sign language users and non-sign language users.
5. Evaluating the system's performance in real-world scenarios to validate its practicality and scalability.

CMRCET B. Tech (AIML) PageNo 2

GesturePath:Real-Time Sign Language Detection with Action Recognition

### Project Scope and Limitations

#### Scope:

1. The project focuses on the development of a real-time gesture recognition system to aid communication for the deaf and mute.
2. It leverages MediaPipe Holistic for detecting hand, face, and body keypoints to capture detailed gestures.
3. The LSTM neural network enables the accurate recognition of dynamic sign language gestures.
4. The system is designed to be scalable, allowing integration into various platforms, including mobile and desktop applications.
5. Provides a practical and inclusive solution for bridging communication gaps between sign language users and others.

**Limitations:**

1. Limited to predefined gesture datasets, requiring additional training for expanding gesture vocabulary.
2. Performance depends on the quality and lighting conditions of the input video feed.
3. The system may face challenges in recognizing overlapping or ambiguous gestures.
4. Computational resource requirements for real-time processing may limit deployment on low-end devices.
5. Accuracy could be affected by variations in individual gesture styles or incomplete gestures.

CMRCET B. Tech (AIML) PageNo 3

**CHAPTER 2**

**BACKGROUND WORK**

CMRCET B.Tech (AIML) PageNo 4

GesturePath:Real-Time Sign Language Detection with Action Recognition

**CHAPTER 2**

**BACKGROUND WORK**

**2.1 Existing Method 1:** **Static Image-Based Recognition**

**2.1.1 Introduction:**

Specialized devices, such as the Leap Motion controller, have revolutionized user interaction by employing advanced technology like infrared sensors to capture intricate hand gestures. These devices enable precise 3D motion tracking, particularly in gaming and design realms, offering a unique hands-on experience. By harnessing the power of infrared technology, these devices detect and interpret movements with impressive accuracy. However, their limitations include restricted gesture range and susceptibility to occlusions, posing challenges in delivering consistent performance across various applications and scenarios

##### 2.1.2. Merits, Demerits, and Challenges

* **Merits**:
  + Enhanced Precision: Specialized devices offer remarkable precision in capturing and interpreting gestures, particularly in 3D space
  + Immersive Experiences: They facilitate immersive experiences in gaming, design, and augmented reality (AR) applications, enhancing user engagement.
  + Advanced Technology: Leveraging infrared sensors and sophisticated technology, these devices enable high-resolution gesture tracking.
* **Demerits**:
  + Limited Gesture Range: These devices often have a confined range for recognizing gestures, limiting their usability in broader spatial environments.
  + Occlusion Sensitivity: Occlusions or interruptions in the line of sight might accurate gesture recognition, affecting overall performance.

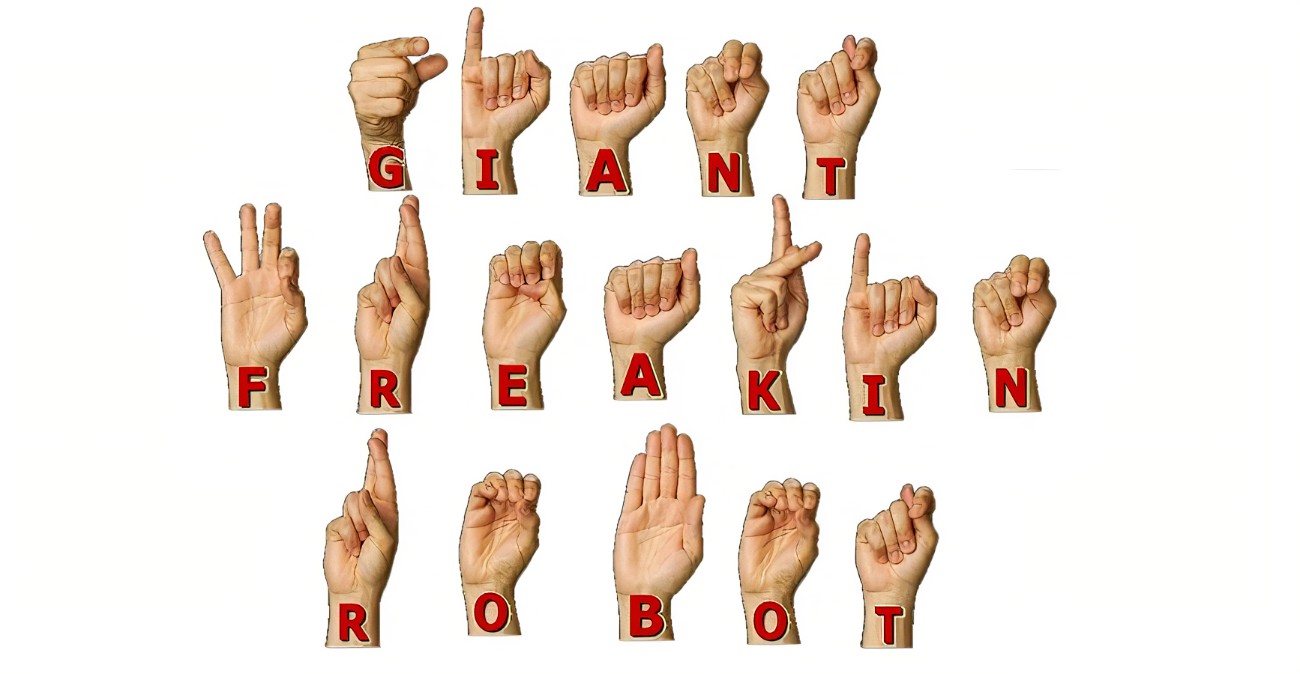
CMRCET B. Tech (AIML) PageNo 5

GesturePath:Real-Time Sign Language Detection with Action Recognition

* **Challenges:**
* Consistency Across Applications: Ensuring consistent performance and accuracy across different software and application domains remains a challenge.
* Interoperability Concerns: Integrating specialized devices with existing technologies or platforms can pose interoperability challenges.
* User Adaptability: Users may face a learning curve in understanding and utilizing the specific functionalities of these

##### 2.1.3. Implementation of Existing Method 1

Integrating specialized devices like the Leap Motion controller involves embedding infrared sensors for intricate hand gesture capture. The system requires calibrated3Dtracking algorithms for precise recognition in gaming, design, and augmented reality applications. However, addressing gesture range limitations necessitates sensor optimization, while addressing occlusion sensitivity requires sophisticated algorithms for uninterrupted tracking, thereby enhancing overall usability and adaptability across diverse applications.



**Figure.2.1: Static Image-Based Recognition**

CMRCET B. Tech (AIML) PageNo 6

GesturePath:Real-Time Sign Language Detection with Action Recognition

### ****2.2. Existing Method 2:**** Dynamic Gesture Recognition Using Hidden Markov Models (HMMs)

### ****2.2.1. Introduction****

This method involves analyzing static images of hand gestures to recognize sign language. Techniques like Histogram of Oriented Gradients (HOG) or Convolutional Neural Networks (CNNs) are commonly used for feature extraction and classification. Utilizing written language, emojis, and multimedia sharing, text-based communication empowers the deaf and mute to engage in real-time conversations, thus enhancing their social inclusion.

#### ****2.2.2. Merits, Demerits, and Challenges****

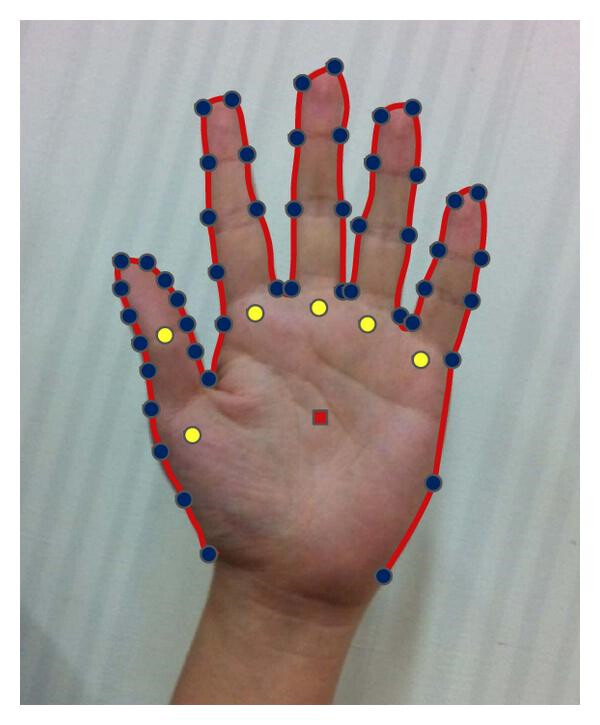
* **Merits**:
  + High accuracy for static gestures.
  + Simplified implementation for single-frame analysis.
* **Demerits**:
  + Lack of Non-Verbal Cues: Absence of non-verbal cues, such as body language and tone, can lead to ambiguity or misunderstanding.
  + Ineffective for dynamic gestures requiring temporal context.
* **Challenges**:
  + Limited scalability to complex or large gesture vocabularies.
  + Requires large, annotated datasets for training.
  + Misinterpretation: Ambiguity in text can lead to misinterpretation.

CMRCET B. Tech (AIML) PageNo 7

GesturePath:Real-Time Sign Language Detection with Action Recognition

#### ****2.2.3. Implementation of Existing Method 2****

* Images of hand gestures are preprocessed to detect hand regions.
* Feature extraction is performed using HOG or CNNs.
* A classifier (e.g., Support Vector Machines or Softmax) is used for gesture recognition.
* However, addressing gesture range limitations necessitates sensor optimization



### Figure.2.2: Dynamic Gesture Recognition

CMRCET B. Tech (AIML) PageNo 8

GesturePath:Real-Time Sign Language Detection with Action Recognition

**2.3. Existing Solution 3: Deep Learning-Based Sign Language Recognition Using CNN**

**2.3.1 Introduction:**

This hybrid approach combines the spatial feature extraction capabilities of CNNs with the temporal sequence modeling. It is a state-of-the-art technique for recognizing dynamic sign language gestures. This approach leverages **Convolutional Neural Networks (CNNs)** for spatial feature extraction and combines it with temporal sequence modeling networks like **LSTMs (Long Short-Term Memory)** or **GRUs (Gated Recurrent Units)** to handle dynamic sign language gestures. CNNs are excellent at analyzing spatial information in image data, such as detecting hand shapes, positions, and orientations. Meanwhile, LSTMs excel at capturing the temporal dependencies between gestures in a sequence, making this hybrid model ideal for recognizing gestures that unfold over time.

#### ****2.2.2. Merits, Demerits, and Challenges****

**Merits:**

* High accuracy for dynamic and complex gestures.
* Scalable to large datasets with diverse gesture vocabularies.
* Real-Time Capability; With optimized hardware, it can perform real-time recognition, making it practical for live applications.

**Demerits:**

* Requires significant computational resources for training.
* Data Dependency; Requires large, high-quality labeled datasets for effective training, which can be difficult to obtain for sign languages with limited resources.
* Risk of overfitting with insufficient data.

**Challenges:**

* Dataset Limitations,Sign languages differ across regions, requiring datasets tailored to specific languages or dialects.
* Gesture Ambiguity; Some gestures may look similar, leading to misclassification.

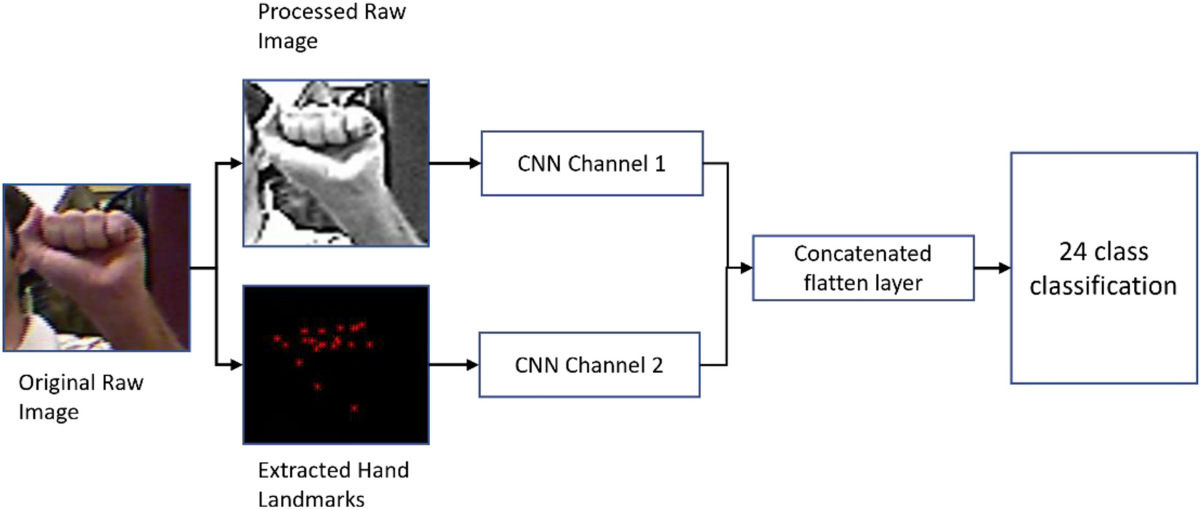
CMRCET B. Tech (AIML) PageNo 9

GesturePath:Real-Time Sign Language Detection with Action Recognition

#### ****2.2.3. Implementation of Existing Method 3****

**Implementation:**

* Preprocessing:Frames are extracted from videos and resized to a consistent input size.Data augmentation techniques (rotation, scaling, noise addition) are used to improve robustn
* CNNs extract spatial features, while LSTMs process temporal dependencies.
* The combined model outputs gesture classifications for text translation.
* Training: Loss functions like categorical cross-entropy are used.
* Deployment: Optimized using techniques like quantization or pruning to make the model lightweight for mobile or edge devices.



### Figure.2.3: Deep Learning-Based Sign Language Recognition

CMRCET B. Tech (AIML) PageNo 10

GesturePath:Real-Time Sign Language Detection with Action Recognition

**CHAPTER 3**

**RESULTS AND DISCUSSION**

CMRCET B. Tech (AIML) PageNo 11

GesturePath:Real-Time Sign Language Detection with Action Recognition

**CHAPTER 3**

**RESULTS AND DISCUSSION**

### ****3.1. Comparison of Existing Solutions****

|  |  |  |  |
| --- | --- | --- | --- |
| **Aspect** | **Static Image-Based Recognition** | **Dynamic Gesture Recognition (HMMs)** | **Deep Learning**  **(CNN)** |
| **Purpose** | Recognizing static hand gestures | Recognizing temporal gesture sequences | Recognizing dynamic gestures with spatial-temporal features |
| **Methodology** | Feature extraction using HOG/CNNs | Temporal modeling with HMMs | CNN for spatial features for temporal dependencies |
| **Strengths** | High accuracy for static gestures | Effective for time-series gestures | High accuracy and scalability for dynamic gestures |
| **Weaknesses** | Ineffective for dynamic gestures | Limited accuracy and scalability | High computational resource requirements |
| **Primary Use Cases** | Simple gesture vocabularies | Sequential gesture recognition in controlled settings | Complex and dynamic gesture vocabularies |

CMRCET B. Tech (AIML) PageNo 12

GesturePath:Real-Time Sign Language Detection with Action Recognition

#### ****3.2. Data Collection and Performance Metrics****

**Data Collection:**

* **Static Image-Based Sign Language Recognition:** Static hand gesture images from open repositories.
* **Dynamic Gesture Recognition Using Hidden Markov Models (HMMs)**: Motion-tracking datasets with labeled dynamic gestures.
* **Deep Learning-Based Sign Language Recognition Using CNN:** Dynamic gesture datasets with annotated keypoints.

#### ****Performance Metrics:****

|  |  |  |  |
| --- | --- | --- | --- |
| **Metric** | **Static Image-Based Recognition** | **Dynamic Gesture Recognition (HMMs)** | **Deep Learning (CNN)** |
| **Accuracy** | ~85% | ~70% | ~90% |
| **Precision** | High for static gestures | Moderate | High |
| **Recall** | Moderate | Low | High |
| **F1-Score** | Moderate | Moderate | High |
| **Inference Time** | Low | Moderate | High |
| **Dataset Dependency** | Moderate | Low | High |

CMRCET B. Tech (AIML) PageNo 13

GesturePath:Real-Time Sign Language Detection with Action Recognition

CHAPTER 4

**CONCLUSION**

CMRCET B. Tech (AIML) PageNo 14

GesturePath:Real-Time Sign Language Detection with Action Recognition

**CHAPTER 4**

**CONCLUSION**

The analysis of the three techniques reveals that each method has its own strengths and specific use cases, balancing performance, security, and data capacity in unique ways.

**Key Findings:**

**Static Image-Based Recognition** is effective for simple and static gestures but fails to handle dynamic and complex gesture vocabularies. **Dynamic Gesture Recognition using HMMs** provides better results for temporal data but suffers from scalability and accuracy challenges in real-world scenarios. **Deep Learning-Based Recognition (CNN)** emerges as the most accurate and scalable solution for dynamic gesture recognition, handling complex gestures efficiently.

* **Performance vs. Security:**

Static Image-Based Methods are low in computational complexity and security risks due to minimal data dependency. **HMM-Based Methods** offer moderate performance but are vulnerable to noise and data inaccuracies. **CNN-LSTM Models** ensure high performance but come with increased computational requirements and potential risks related to large dataset handling.

* **Data Capacity:**

**Static Image-Based Recognition** operates efficiently on small datasets. **HMMs** require moderate-sized datasets with well-labeled temporal data. **CNN** demands large, annotated datasets, which can strain data storage and processing resources.

In conclusion, while each method has its unique strengths and limitations, the CNN-LSTM approach proves to be the most effective solution for real-time dynamic sign language recognition. Future work may focus on optimizing computational efficiency, reducing dataset dependency, and expanding vocabulary coverage for broader application scenarios.

CMRCET B. Tech (AIML) PageNo 15

GesturePath:Real-Time Sign Language Detection with Action Recognition

**CHAPTER 5**

**REFERENCES**

CMRCET B. Tech (AIML) PageNo 16

GesturePath:Real-Time Sign Language Detection with Action Recognition

**CHAPTER 5**

## REFERENCES

1. Ravishankar, V., & Kishore, M. (2018). Static Hand Gesture Recognition for Sign Language Interpretation. International Journal of Computer Applications, 179(7), 31-36.
2. Zhou, Z., & Liu, T. (2014). Static Hand Gesture Recognition Using Convolutional Neural Networks. Proceedings of the 8th International Conference on Image and Graphics.
3. Kumar, P., & Anand, R. (2015). Real-Time Indian Sign Language Recognition Using Static Hand Gestures. Proceedings of the International Conference on Electrical, Electronics, and Optimization Techniques.
4. Starner, T., & Pentland, A. (1995). Visual Recognition of American Sign Language Using Hidden Markov Models. Proceedings of the International Workshop on Automatic Face- and Gesture-Recognition.
5. Huang, G., & Shi, J. (2006). Gesture Recognition with Hidden Markov Models. Proceedings of the International Symposium on Visual Computing.
6. Parmar, V., & Shah, D. (2018). Sign Language Recognition Using Hidden Markov Models. International Journal of Computer Science and Information Security, 16(7), 51-55
7. Girdhar, R., & Ramanan, D. (2017). Recognizing Sign Language Using Deep Learning. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(8), 1475-1487.
8. Aytar, Y., & Zisserman, A. (2015). The 2015 Kinetics Dataset. Proceedings of the IEEE International Conference on Computer Vision (ICCV).
9. Yu, L., & Li, H. (2018). A Deep Learning Approach for Real-Time Sign Language Recognition Using CNN-LSTM. Proceedings of the International Conference on Artificial Intelligence and Computer Vision (AICV).
10. Wang, H., & Yang, L. (2020). Sign Language Recognition Using Hybrid CNN-LSTM Models. Computer Vision and Image Understanding, 198, 102984.

CMRCET B. Tech(AIML) PageNo 17