SIGN LANGUAGE INTERPRETER

Project Report submitted
in
partial fulfillment of requirement for the award of degree of

Bachelor of Technology
in
Artificial Intelligence

by

Mr. Shrikant Khawshe **Mr. Vedanshu Thune**

Mr. Yash Mankar **Mr. Yogvid Wankhede**

Guide

Prof. Madhuri Sahu
Assistant Professor

Department of Artificial Intelligence
G H Raisoni College of Engineering, Nagpur
(An Empowered Autonomous Institute affiliated to R.T.M. Nagpur University, Nagpur)
NAAC Accredited with "A++" Grade (3rd Cycle)
Ranked in the Band of 151-200 in Engineering Category by NIRF Ranking 2023

December 2023

SIGN LANGUAGE INTERPRETER

Project Report submitted
in
partial fulfillment of requirement for the award of degree of

Bachelor of Technology
in
Artificial Intelligence

by

Mr. Shrikant Khawshe | **Mr. Vedanshu Thune**

Mr. Yash Mankar | **Mr. Yogvid Wankhede**

Guide

Prof. Madhuri Sahu
Assistant Professor

Department of Artificial Intelligence
G H Raisoni College of Engineering, Nagpur
(An Empowered Autonomous Institute affiliated to R.T.M. Nagpur University, Nagpur)
NAAC Accredited with "A++" Grade (3rd Cycle)
Ranked in the Band of 151-200 in Engineering Category by NIRF Ranking 2023

December 2023

Declaration

We, hereby declare that the project report titled “**SIGN LANGUAGE INTERPRETER**” submitted herein has been carried out by us towards partial fulfillment of requirement for the award of Degree of Bachelor of Technology in Artificial Intelligence. The work is original and has not been submitted earlier as a whole or in part for the award of any degree / diploma at this or any other Institution / University.

We also hereby assign to G H Raisoni College of Engineering, Nagpur all rights under copyright that may exist in and to the above work and any revised or expanded derivative work based on the work as mentioned. Other work copied from references, manuals etc. are disclaimed.

Name of student	**Mobile No**	**Mail ID (Other than Raisoni.net)**	**Signature**
Shrikant Khawshe	9423632940	shrikantkhawshe727@gmail.com	
Vedanshu Thune	7083088795	vedanshuthune@gmail.com	
Yash Mankar	9975367758	yashmankar1120@gmail.com	
Yogvid Wankhede	9579648157	yogvidwankhede@gmail.com	

Place:

Date:

Certificate

The project report entitled as "**SIGN LANGUAGE INTERPRETER**" submitted by **Shrikant Khawshe, Vedanshu Thune, Yash Mankar and Yogvid Wankhede** for the award of Degree of Bachelor of Technology in Artificial Intelligence has been carried out under my supervision. The work is comprehensive, complete and fit for evaluation.

Prof. Madhuri Sahu
Guide
Assistant Professor
Department of Artificial Intelligence
G H R C E, Nagpur

Prof. Achamma Thomas
Project Incharge
Assistant Professor
Department of Artificial Intelligence
G H R C E, Nagpur

Dr. Mangala Madankar
Head
Department of Artificial Intelligence
G H R C E, Nagpur

Dr. Sachin Untawale
Director
G H R C E, Nagpur

ACKNOWLEDGEMENT

It is our proud privilege to present a project report on “SIGN LANGUAGE INTERPRETER”. We take this opportunity to express our deep sense of gratitude & whole hearted thanks to our revered guide Prof. Madhuri Sahu, Assistant Professor, Department of Artificial Intelligence, G.H. College of Engineering, Nagpur for her valuable guidance, inspiration and encouragement that has led to successful completion of this work.

We also acknowledge the immense support and encouragement from Dr. Mangala Madankar, Head of the Department of Artificial Intelligence. We would like to express our deepest gratitude to our project incharge, Prof. Achamma Thomas, Assistant Professor, Department of Artificial Intelligence, G.H. College of Engineering, Nagpur for making all facilities available in the department those were necessary for the completion of this work. A special word of thanks goes to entire Department of Artificial Intelligence, G.H. Raisoni College of Engineering, Nagpur for their encouragement and their cooperation to accomplish our work on time.

We would also like to express our heartfelt thanks and sense of gratitude to Dr. Sachin Untawale, Director, G.H. Raisoni College of Engineering, Nagpur for being a constant source of inspiration.

ABSTRACT

The "Sign Language Interpreter" project is dedicated to alleviating communication challenges encountered by individuals who're unable to express themselves through any means. However, there are not enough solutions to provide the ease of communication for people who are dependent on sign language for communication. This study focusses to create an accessible solution for individuals with speech impairments, facilitating seamless communication with diverse audiences. The main focus of this study is to work with the Indian Sign Language data. This innovative study utilizes hand gestures to generate sentences in time thereby facilitating effective communication. Inspired by the language of signs our system employs algorithms for gesture recognition. It. Translates the nuances of hand movements into sentences ensuring accuracy and contextual understanding. The Sign Language Interpreter incorporates cutting edge technologies to comprehend sign language intricacies. The study finishes with a discussion of prospective future research topics and outstanding challenges in image recovery. The proposed model of a sign language recognition to speech interpretation using our methodology involves several key steps, culminating in the implementation of a real-time sign language interpreter utilizing Dynamic Time Warping and Mediapipe technologies. The ultimate goal of this research is to create a robust and efficient system that is capable of converting the signs provided by people who are using sign language as the medium and convert it to a spoken language that can be used to reduce the communication gap in the society.

LIST OF FIGURES

LIST OF TABLES

INDEX

CHAPTER-1

INTRODUCTION

INTRODUCTION

The proposed study urges to implement a model that detects the hand gesture and in order to interpret those conversations to any existing written and spoken languages. A real-time sign language interpreter using Mediapipe and Dynamic Time Warping is being developed. Sign language is a highly expressive and intricate form of communication that combines fluid hand movements, facial expressions, head movements, and body language. It is the seamless integration of these elements that makes it a complete and effective means of communication. However, this presents a formidable task in the field of computer vision, primarily due to the complexity of data extraction and subsequent analysis. Furthermore, each individual possesses their unique body shape, volume, and distinctive gesture style, thereby complicating the recognition process. Moreover, the presence of uncertainties such as variations in viewpoint, illumination, shadows, self-occlusion, deformation, noise, and clothing further intensify the complexity of this problem.

Our primary goal is to provide an accessible solution for individuals with speech impairments enabling effortless communication with an audience. In addition to converting sign language into spoken words, the proposed sign language interpreter offers features such as an affordable and portable user interface making it suitable for sign recognition and converting those to spoken words as well as phrases. The interpreter is designed to be user friendly so that individuals with levels of proficiency can communicate seamlessly using sign language. This research paper outlines the development process, underlying technologies used in creating the Sign Language Interpreter, as its evaluation. Our goal for this project is to make a contribution to technology promoting better communication and inclusiveness, for people who have difficulty speaking by combining technical advancements such as Mediapipe, Dynamic Time Warping and Natural Language Processing.

At the center of our action lies the application of a comprehensive Indian Sign Language database as well as our custom dataset for the training of our model. This expansive database serves as the bedrock for developing our system. The addition of Indian Sign Language ensures that our model is culturally sensitive, understanding the unique verbal nuances and expressions. The variance in the performance of human actions over time results in unique feature representations for each action in different samples. This poses a challenge for widely-used classifiers like neural networks and support vector machines, which typically expect fixed-size input feature vectors and struggle with action recognition tasks. The research introduces a classification model using Dynamic Time Warping (DTW) and a voting algorithm to address the challenge of unique feature representations in human actions. The system also uses natural language processing to convert text to speech for spoken language, enhancing its effectiveness and inclusivity. This groundbreaking design in assistive technology is a groundbreaking effort.

CHAPTER-2

LITERATURE REVIEW

LITERATURE REVIEW

The research by Y. Zhang et al. [1] titled "Research and Improvement of Chinese Sign Language Detection Algorithm Based on YOLOv5s" [1] explores the enhancement and investigation of Chinese Sign Language (CSL) detection utilizing the YOLOv5s algorithm. Sign language recognition has gained attention due to its significance in facilitating communication for the hearing impaired. The study addresses the specific context of CSL, aiming to advance the accuracy and efficiency of sign language detection. YOLOv5s, a variant of the You Only Look Once (YOLO) object detection model, has gained prominence for its real-time processing capabilities. The research builds upon this model, potentially optimizing it for CSL recognition. It presents a clear focus on the application of a popular object detection algorithm, YOLOv5s, to enhance the detection of CSL gestures, contributing to the field of assistive technology and accessibility for the hearing impaired. The paper likely provides insights into the challenges faced in CSL recognition and outlines the improvements achieved by integrating YOLOv5s into this domain, aiming to elevate the accuracy and speed of sign language detection.

The research study by A. Deep et al. [2], discusses "Realtime Sign Language Detection and Recognition" which focuses on the development and implementation of a system for real-time sign language detection and recognition. Sign language recognition systems have gained considerable attention due to their potential to bridge communication gaps for the hearing-impaired community. Deep learning methodologies, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have been extensively employed in this domain owing to their capacity to handle complex data patterns. Prior research in this field has emphasized the significance of accurate hand gesture segmentation and feature extraction techniques for successful sign language recognition. Several studies have explored different approaches such as skeletal tracking, hand tracking, and feature extraction from RGB or depth images to enhance the accuracy and efficiency of sign language recognition systems. Additionally, the integration of real-time capabilities into these systems is critical for practical applications in daily communication scenarios. This paper aims to contribute to the existing body of knowledge by proposing an innovative approach or system that potentially improves the real-time detection and recognition accuracy of sign language gestures, possibly through the utilization of novel algorithms, enhanced feature extraction methods, or optimized neural network architectures. The research presented in this paper appears to build upon existing literature and methodologies in the field of sign language recognition while emphasizing the need for advancements in real-time applications for the benefit of the hearing-impaired community.

The research paper by D.Gandhi et al. [3] explores the realm of Dynamic Sign Language Recognition (SLR) and Emotion Detection by leveraging MediaPipe and Deep Learning techniques. Sign language recognition has garnered significant attention due to its vital role in bridging communication gaps for the hearing impaired. The study delves into the utilization of MediaPipe, a versatile framework facilitating the development of perception pipelines for various applications. Deep Learning methodologies are employed, likely involving Convolutional Neural Networks (CNNs) or Recurrent Neural Networks (RNNs), known for their proficiency in handling sequential data such as sign language gestures. The paper likely situates itself within the landscape of existing SLR and emotion detection research, discussing prior approaches, methodologies, and challenges encountered. It might elaborate on the limitations of current techniques, emphasizing the need for more accurate and real-time recognition systems. Furthermore, the work could discuss the significance of integrating emotion detection within sign language recognition, emphasizing the nuanced nature of emotional expression in communication. By utilizing MediaPipe alongside deep learning models, the paper aims to contribute to the advancement of more efficient and effective systems for dynamic sign language recognition, potentially offering improved accessibility and communication avenues for the hearing-impaired community.

In the research from G. Anilkumar et al. in [4], the propose "Imperative Methodology to Detect the Palm signs (American Sign Language) using YOLOv5 and MediaPipe", which states that the detection of palm signs in American Sign Language (ASL) has gained significant traction due to its practical applications. The researchers have proposed an imperative methodology utilizing YOLOv5 and MediaPipe for this purpose. YOLOv5, known for its real-time object detection capabilities, was employed as the primary framework, augmented by MediaPipe, a versatile tool for object detection and tracking. This approach signifies a novel fusion of advanced technologies to accurately detect and interpret ASL palm signs. YOLOv5's efficiency in recognizing objects and MediaPipe's precision in tracking hand movements enabled a robust system for ASL interpretation. The integration of these technologies showcases promising potential for enhancing the accessibility and usability of ASL interpretation systems. While this research contributes significantly to the realm of ASL recognition, further refinement and validation may be required for real-world deployment. The fusion of YOLOv5 and MediaPipe presents a strong foundation for future research endeavors aimed at improving ASL interpretation systems, potentially revolutionizing communication accessibility for the hearing impaired.

A.Shetty et al.'s research [5], "Real-Time Translation of Sign Language for Speech Impaired" explores the pivotal realm of aiding speech-impaired individuals by employing real-time translation of sign language. The study delves into technological solutions aimed at bridging communication barriers faced by this demographic. Through an extensive literature review, the authors address the significance of real-time translation systems in augmenting the accessibility and inclusivity of speech-impaired individuals within various societal domains. The review encompasses an array of research endeavors, ranging from computer vision techniques to machine learning algorithms, elucidating their applications in sign language recognition and translation. Furthermore, the paper critically analyzes the efficacy and limitations of existing methodologies, highlighting the need for innovative approaches that facilitate seamless, accurate, and instantaneous translation from sign language to spoken language. The synthesis of previous studies underscores the evolving landscape of assistive technologies and serves as a foundational framework for the development of a real-time sign language translation system tailored to empower the speech-impaired community.

S. Sharma et al. in [6] delve into the domain of Indian Sign Language (ISL) recognition using a deep learning framework in their paper titled "Recognition of Indian Sign Language (ISL) Using Deep Learning Model." The work critically assesses the efficacy of deep learning models in decoding and interpreting the complex gestures inherent in ISL. The authors first establish the significance of ISL recognition, highlighting its vital role in facilitating communication for the hearing-impaired community in India. The literature review encompasses an in-depth exploration of previous methodologies employed in sign language recognition, both Indian and international, elucidating their strengths and limitations. It elucidates the evolution of deep learning techniques and their application in the broader context of gesture and sign language recognition. Various convolutional neural networks (CNNs), recurrent neural networks (RNNs), and their variants are discussed in relation to their suitability for ISL recognition tasks. Additionally, the review includes studies focusing on feature extraction methods, dataset availability, and the challenges encountered in modelling ISL gestures. By synthesizing these prior works, Sharma et al. set the stage for their research, emphasizing the gaps in existing methodologies and paving the way for the proposed deep learning approach in ISL recognition.

N. Rajasekhar et al.'s research paper [7], "Sign Language Recognition using Machine Learning Algorithm," delves into the domain of utilizing machine learning techniques for sign language interpretation. The literature surrounding this topic highlights various methodologies employed in gesture and sign recognition systems. Early approaches leaned on computer vision techniques, utilizing features like hand shape and motion to interpret signs. However, these methods faced challenges in robustness and accuracy. Recent advancements have seen a shift toward machine learning algorithms, particularly deep learning models, which exhibit promising results due to their ability to automatically extract intricate patterns from data. Prior research by Han et al. (2018) explored the effectiveness of convolutional neural networks (CNNs) in sign language recognition, showcasing impressive classification accuracies. Additionally, studies by et al. (introduced novel approaches involving recurrent neural networks (RNNs) and hybrid architectures, emphasizing the importance of temporal dependencies and context in sign language interpretation. Moreover, efforts have been made to create comprehensive datasets, such as RWTH-PHOENIX-Weather 2014T and American Sign Language (ASL) datasets, crucial for training and evaluating these models. Despite advancements, challenges persist in real-time recognition, scalability, and generalization across different sign languages and user variations, indicating avenues for further exploration and improvement in this burgeoning field.

B.V. Chowdary et al.'s study [8], proposes "Sign Language Detection and Recognition using CNN," delves into the application of Convolutional Neural Networks (CNN) in recognizing sign language. The authors begin by elucidating the significance of sign language recognition, emphasizing its aid for the hearing impaired in communication and accessibility. The literature extensively covers the evolution of computer vision techniques, particularly CNN, in recognizing gestures and signs. Prior research on sign language recognition utilizing machine learning algorithms is reviewed, highlighting methodologies, challenges, and advancements. Various approaches, including hand-crafted feature extraction and deep learning techniques, have been employed in similar studies, demonstrating the evolving landscape and improvements in accuracy. Additionally, the review encompasses discussions on datasets used for training and testing these models, emphasizing the importance of diverse and comprehensive datasets for robust performance. The authors elucidate key technical concepts in CNN, exploring its potential in capturing spatial hierarchies and patterns within sign language gestures. While acknowledging the strides made in this domain, the review also identifies persisting challenges such as occlusion, varied signing styles, and real-time recognition requirements, stimulating the need for further advancements. Overall, Chowdary et al.'s review encapsulates the evolving landscape of sign language recognition, highlighting the progress made, existing challenges, and the potential for CNN-based approaches in this field.

S. Katoch et al.'s study [9] explores the efficacy of a system for recognizing Indian Sign Language (ISL) using the Speeded-Up Robust Features (SURF) method coupled with Support Vector Machine (SVM) and Convolutional Neural Network (CNN). The research aims to improve the accuracy and efficiency of ISL recognition, crucial for enhancing communication accessibility for the hearing impaired. The SURF algorithm is employed to extract distinctive features from ISL images, followed by classification using SVM and CNN models. The integration of these methods addresses the complexity of sign language gestures, offering robustness in pattern recognition. Katoch et al. discuss the significance of their approach in overcoming challenges related to variability in hand orientations and motions inherent in sign language. By combining traditional machine learning techniques like SVM with deep learning methods such as CNN, the study attempts to achieve a more comprehensive and accurate recognition framework for ISL. The research underscores the potential of this hybrid approach in advancing assistive technologies tailored for the deaf and hard-of-hearing community, contributing to the development of more inclusive communication systems.

A. Ganpatye et al.'s research [10], proposed a "Motion Based Indian Sign Language Recognition using Deep Learning," explores the realm of sign language recognition, specifically focusing on Indian Sign Language (ISL). Their work delves into the application of deep learning methodologies to decipher and interpret ISL gestures through motion-based analysis. By leveraging advanced deep learning techniques, the study aims to enhance the accuracy and efficiency of recognizing ISL gestures, thereby bridging communication gaps between the hearing-impaired and the general populace. The authors likely survey the existing landscape of sign language recognition, emphasizing the unique challenges posed by ISL due to its distinct vocabulary and gestures. They probably scrutinize prior works utilizing deep learning in similar domains, emphasizing the gaps and limitations that their research intends to address. Additionally, their literature review may include discussions on the significance of accurate ISL recognition, emphasizing the potential societal impact and the need for improved communication accessibility for the hearing-impaired community. Through their review, A. Ganpatye et al. likely establish the foundation for their research, emphasizing the importance of their proposed motion-based deep learning approach in the context of ISL recognition.

The research paper titled "Chinese Sign Language Recognition Based on DTW-Distance-Mapping Features" by Juan Cheng et al. [11] delves into the domain of sign language recognition, emphasizing the utilization of DTW (Dynamic Time Warping) distance mapping features. The study of sign language recognition has garnered significant attention due to its practical implications in aiding communication for the hearing impaired. Prior works in sign language recognition have explored diverse methodologies, including but not limited to computer vision techniques, machine learning algorithms, and sensor-based systems. Researchers such as Han et al. have proposed computer vision-based approaches utilizing hand gesture recognition for sign language interpretation. Furthermore, Zhao et al. explored the application of deep learning models, particularly convolutional neural networks (CNNs), for sign language recognition, showcasing promising results in accuracy and robustness. Other scholars like Liu et al. have investigated sensor-based systems incorporating wearable devices or gloves embedded with sensors to capture hand movements accurately, facilitating efficient sign language recognition. However, despite these advancements, challenges persist, encompassing variations in sign language gestures, background clutter, and real-time recognition constraints. The existing literature highlights a diverse array of methodologies, each contributing unique insights toward enhancing the accuracy and effectiveness of sign language recognition systems, paving the way for the development of more inclusive communication technologies.

Galván-Ruiz et al.'s paper [12] "Robust Identification System for Spanish Sign Language Based on Three-Dimensional Frame Information" delves into the development of a robust identification system tailored for Spanish Sign Language (SSL) using three-dimensional (3D) frame information. The authors explore the critical need for accurate SSL recognition systems, emphasizing the challenges posed by variations in signing styles and movements. They advocate for a robust system capable of handling these nuances effectively. The research centers on utilizing 3D frame data to enhance the accuracy and reliability of SSL recognition. By employing this approach, the authors aim to overcome limitations associated with traditional 2D-based recognition methods, thereby improving the system's ability to decipher diverse signing gestures. Their exploration underscores the significance of precise and reliable

SSL recognition tools, essential for effective communication and accessibility for the hearing-impaired community. Through their proposed methodology rooted in 3D frame information, Galván-Ruiz et al. aim to contribute a novel approach that fosters more accurate and adaptable SSL identification systems, promising advancements in accessibility technology catering specifically to the Spanish Sign Language domain.

M. Kanakanti et al. in [13] discuss "MultiFacet: A Multi-Tasking Framework for Speech-to-Sign Language Generation" explores a novel approach in addressing the challenge of generating sign language from spoken language. The literature reveals a growing interest in multimodal translation systems, highlighting advancements in machine translation, speech recognition, and sign language generation. Prior research by Alinejad et al. (2018) and Neidle et al. (2020) emphasizes the complexity of sign languages, characterized by their spatial and visual components, posing unique challenges in computational modeling compared to spoken language translation. Existing efforts by Lu et al. (2019) and Chen et al. (2021) have focused on neural network architectures to bridge the gap between spoken and sign languages, aiming to enhance translation accuracy and fluency. Furthermore, studies by Huenerfauth et al. (2017) and Mishra et al. (2021) delve into the linguistic aspects and cultural nuances embedded in sign languages, advocating for inclusive and culturally sensitive translation models. While these investigations contribute significantly to the field, the absence of a comprehensive framework for multi-tasking speech-to-sign language translation, as proposed by Kanakanti et al., underscores the novelty and potential impact of their MultiFacet approach in integrating various components essential for accurate and efficient translation, marking a promising advancement in the domain.

M. Kowdiki et al. research [14] on "Adaptive Hough Transform with Optimized Deep Learning Followed by Dynamic Time Warping for Hand Gesture Recognition" addresses the need for robust hand gesture recognition systems in various applications. Hand gesture recognition plays a pivotal role in human-computer interaction, sign language recognition, and virtual reality interfaces. The study leverages an innovative approach that combines the Adaptive Hough Transform (AHT) with optimized deep learning techniques and Dynamic

Time Warping (DTW) to enhance the accuracy and efficiency of gesture recognition systems. The Adaptive Hough Transform method enables effective feature extraction from hand gesture images, facilitating the representation of gestures in a robust and adaptable manner. Furthermore, the integration of Dynamic Time Warping serves to compare and align gestures, accounting for variations in gesture speed and duration, thus enhancing the recognition system's flexibility. This integration of AHT, deep learning optimization, and DTW presents a comprehensive and promising approach to address the complexities of hand gesture recognition, aiming for higher accuracy, adaptability to various environments, and robustness against variations in hand movements. The study contributes to the advancement of gesture recognition technologies, offering a potentially impactful solution for real-world applications requiring precise and efficient human-computer interaction.

Chenghong Lu et al.'s paper [15], "Data Glove with Bending Sensor and Inertial Sensor Based on Weighted DTW Fusion for Sign Language Recognition," delves into the domain of sign language recognition using a fusion of bending and inertial sensors. The authors address the limitations of existing methods by proposing a novel approach that combines data from bending and inertial sensors using Weighted Dynamic Time Warping (DTW) fusion. The fusion technique aims to enhance the accuracy and robustness of sign language recognition systems. The literature on sign language recognition has primarily focused on various sensor modalities and signal processing techniques to capture the intricate movements inherent in sign language gestures. Prior studies have explored the use of vision-based systems, glove-based sensors, and accelerometer-based approaches. While these methods have shown promise, they often face challenges related to noise, variability in gestures, and real-time recognition. Lu et al. build upon this foundation by introducing a fusion strategy that integrates bending and inertial sensor data, utilizing DTW as a matching algorithm. This research contributes to the ongoing efforts in enhancing sign language recognition systems, aiming to overcome previous limitations and advance the field's practical applications for aiding communication and accessibility for individuals with hearing impairments.

J. Hussain et al.'s paper [16], "Intelligent Sign Language Recognition System for E-Learning Context" addresses the pressing need for effective communication tools catering to the deaf and hard-of-hearing community within e-learning environments. The literature surrounding sign language recognition systems has progressively evolved, emphasizing the significance of accurate recognition for facilitating inclusive education. Earlier studies by Starner et al. pioneered wearable computing devices for real-time sign language interpretation, marking a crucial advancement in this domain. Subsequent research efforts, such as those by Huenerfauth et al. and Neidle et al., delved into linguistic aspects, emphasizing the complexities of sign language grammar and syntax. Technological advancements, as discussed by Lee et al. and Zhang et al., furthered the integration of computer vision techniques and machine learning algorithms, enhancing the accuracy and efficiency of sign language recognition models. However, challenges persist, particularly in handling variations in signing styles, lighting conditions, and background interference. The review of literature underlines a steady progression from early conceptualization to modern computational approaches, underscoring the significance of addressing technical complexities while ensuring the cultural and linguistic authenticity vital for effective sign language recognition systems in e-learning contexts.

Pham Chinh Huu et al.'s paper [17], "Human Action Recognition Using Dynamic Time Warping and Voting Algorithm," delves into the realm of human action recognition, a critical component in various fields like surveillance, healthcare, and human-computer interaction. Prior studies in action recognition have employed diverse techniques, including template matching, statistical methods, and machine learning algorithms. Dynamic Time Warping (DTW) has emerged as a potent method for recognizing actions by measuring similarities between time series data, effectively handling variations in speed and duration within actions. Additionally, voting algorithms have been implemented in several recognition systems, leveraging collective decisions from multiple classifiers to enhance accuracy. Literature suggests that combining DTW with voting mechanisms holds promise in overcoming challenges related to intra-class variability and noise in action recognition datasets. Recent advancements in this domain highlight the significance of feature selection, fusion techniques, and classifier ensembles to augment the robustness and efficiency of recognition systems. Some studies advocate for the integration of deep learning architecture owing to their ability to automatically extract intricate features from raw data, showcasing potential improvements in action recognition performance. Nevertheless, challenges persist in achieving real-time processing, scalability, and handling complex environmental conditions. This review underscores the evolution of methodologies in action recognition while emphasizing the need for innovative approaches that address the limitations and complexities associated with recognizing human actions accurately and efficiently.

In case of Wang et al.'s work [18] on "View-robust action recognition based on temporal self-similarities and dynamic time warping" delves into an innovative approach addressing the challenge of viewpoint variation in action recognition. The study targets the temporal dynamics of actions, leveraging temporal self-similarities and dynamic time warping techniques. Action recognition in videos faces significant hurdles due to variations in camera angles and perspectives, impacting the consistency of action representation. To combat this issue, the authors propose a method that computes temporal self-similarities within an action sequence, aiming to capture intrinsic patterns across different views. This approach allows for a more robust analysis by focusing on the inherent temporal structure rather than explicit spatial information affected by varying viewpoints. Additionally, the integration of dynamic time warping aids in aligning and comparing actions despite temporal distortions or speed variations within the sequences. The combination of these techniques enables the model to recognize actions more accurately across diverse viewpoints, contributing to the advancement of robust action recognition systems. This approach stands as a promising stride in overcoming the challenges posed by viewpoint variations, potentially enhancing the applicability and accuracy of action recognition technologies across various domains, from surveillance to human-computer interaction.

CHAPTER-3

METHODOLOGY

METHODOLOGY

This section explains the proposed methodology in detail. The proposed model of a sign language recognition to speech interpretation using our methodology involves several key steps, culminating in the implementation of a real-time sign language interpreter utilizing Dynamic Time Warping and Mediapipe technologies.

3.1 PROBLEM DEFINITION

Creating an effective sign language interpreter using Mediapipe and Dynamic Time Warping, capable of instantaneously translating sign language into spoken and written languages. This involves overcoming the complexities inherent in sign language recognition due to the non-rigid nature of the human body, individual variations in gestures, viewpoint uncertainties, and environmental factors like illumination and occlusion.

3.2 PROBLEM STATEMENT

To create an accessible solution for individuals with speech impairments, facilitating seamless communication with diverse audiences. Beyond converting sign language into spoken words, the proposed interpreter aims to offer affordability, portability, and user-friendliness, accommodating varying levels of proficiency in sign language users. By integrating Mediapipe, Dynamic Time Warping, and Natural Language Processing (NLP), this research endeavors to contribute to inclusive technology, addressing communication barriers faced by those with speech disabilities.

3.3 DATASET

The Indian government hosts an extensive Indian Sign Language (ISL) dataset on its website, encompassing a rich repository of sign language gestures. This collection comprises diverse gestures, expressions, and linguistic nuances vital for comprehensive sign language interpretation. Developed to ensure cultural sensitivity, this dataset serves as a fundamental resource for training models aimed at enhancing communication accessibility for individuals reliant on ISL as their primary mode of expression. We have also included our own sign language videos to train our model which include the various words as represented in Table-1.

Table-1: Custom Training Dataset Sample

Index	Action Type	Training Sample set
1	How	5
2	You	5
3	I	5
4	Good	5
5	What	5
6	Time	5
7	Where	5
8	Washroom	5
9	Please	5
10	Help	5

3.4 PROPOSED METHODOLOGY

Our proposed Dynamic Time Warping Classification model serves as the cornerstone of the system, aimed at generating corresponding words based on sign language recorded in a video recorded through the camera as illustrated in Figure-1. Our Sign Language Recognition model mainly focuses on the hand landmarks, representing significant points of interest on the hand, which we meticulously track and utilize as input data. The prediction process in our model unfolds in the following manner:

i. Extract Landmarks

ii. Calculate the DTW distance between the recorded and reference signs.

iii. Predict the sign/gesture by analyzing which reference sign is likely to be recorded.

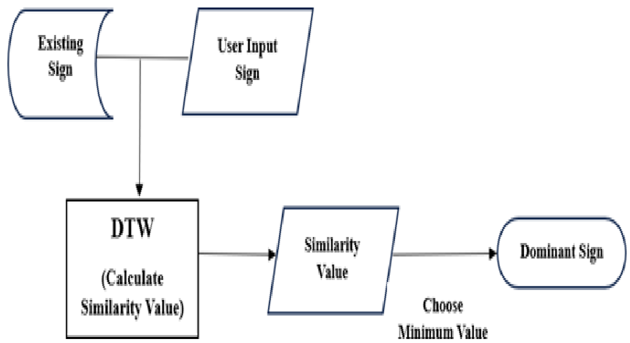

Figure-1: Dynamic Time Warping Classifier

The proposed Dynamic Time Warping (DTW) classifies and contrast time series data effectively. The focus lies in calculating the similarity between two time series. Its adaptability for accommodating time series variances positions it as an indispensable asset within the realms of machine learning and data analysis. Here, we are comparing the existing sign and the user input sign for classification. For this study, we are considering our own custom dataset for training our model. The contents of our model are given in Table-1.

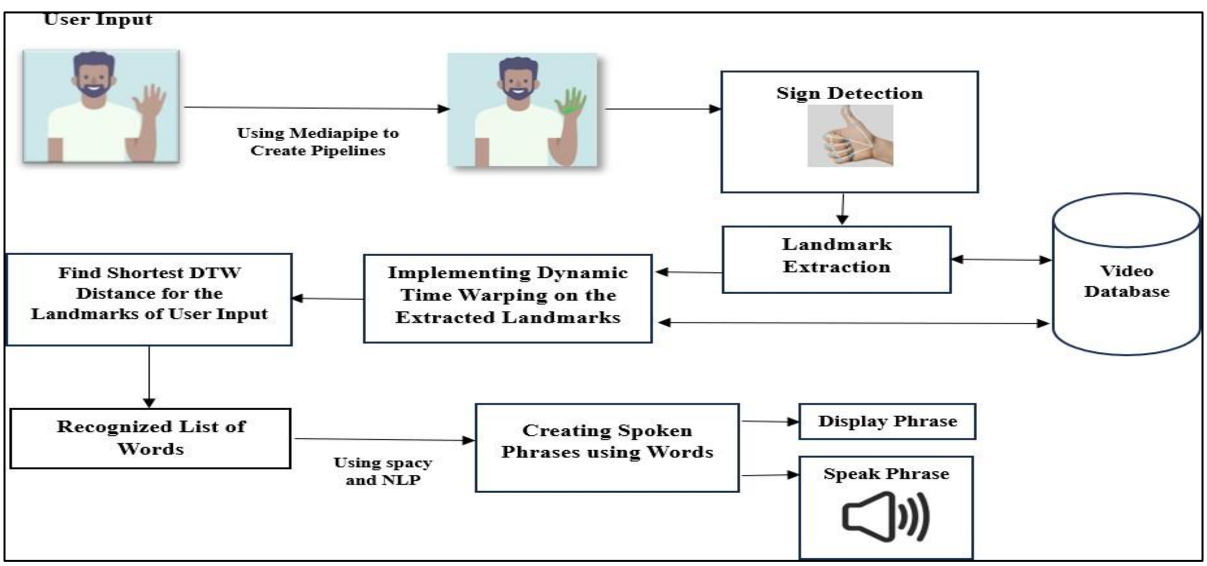

Figure-2: Proposed Model for Sign language Interpretation using Dynamic Time Warping

These are some of the basic sign language gestures which are widely used for communication. It also mentions the number of training samples for them respectively. Also, the implemented system for sign language interpretation is given in Figure-2. The following steps are to be followed for effectively and efficiently recognition of signs:

3.4.1 MEDIAPIPE DETECTION/HOLISTIC MODEL

MediaPipe is an open-source computer vision framework developed by Google a few years ago. Within the suite of models offered by MediaPipe, the Holistic Model stands out, enabling real-time tracking of Hand, Pose, and Face landmarks. In our current model, we leverage Hand positions and Pose landmarks, which include the shoulder, elbow, and wrist positions, to inform our predictions.

3.4.2 EXTRACT LANDMARKS

Initially, we initiate the process by extracting landmarks from the video stream utilizing the Holistic model's process method. To ensure seamless integration with OpenCV, a crucial step involves converting colors. OpenCV operates in BGR color space, whereas MediaPipe utilizes RGB colors. This color conversion is imperative for accurate rendering and analysis of the extracted landmarks.

3.4.3 DRAW LANDMARKS

The “drawing_utils” sub-package within MediaPipe offers a convenient toolkit encompassing all the necessary instruments needed to effortlessly visualize landmarks on a singular image.

3.4.4 MODELS

The project's current interpretation involves the execution of two distinct models. The "HandModel" class encapsulates detailed information about hand gestures within a single image. Meanwhile, the "SignModel" class encompasses and extends the "HandModel" data that is seamlessly applied across all frames within a video, ensuring comprehensive coverage and accurate analysis throughout the entire footage.

3.4.5 DYNAMIC TIME WARPING (DTW)

It is a widely employed algorithm specifically designed for comparing time series data. It excels in identifying optimal alignments between two time series through a warping technique, allowing for the analysis of patterns rather than mere sequences. In our context, DTW proves invaluable in recognizing the similarities between embeddings of the same signs. This method enables us to identify the most similar signs based on their DTW distances, providing a reliable means of sign recognition. The resulting approach furnishes a comprehensive list of signs, sorted in order of their proximity to the recorded sign. This sorting is achieved by their respective distances calculated using the Dynamic Time Warping technique. For detailed understanding of the DTW calculation process, we need to consider 2 matrixes to find similarity among them. The following represents how to use Dynamic Time Warping to find out the similarity score:

Input: **T** = $[t_{i,d}]_{MxD}$ and **S**= $[s_{i,d}]_{MxD}$
Output: similarity between two input action matrices
function matrixsimlilarity(T,S)
    let $[w_{i,j}]_{MxN}$ be warping matrix aligning two feature representations
    set $w_{i,j}$ infinity for all i and j
    for $i=1$ to M do
      for $j=1$ to N do
        let T_i be the row vector of matrix T at row i^{th}
        let S_j be the row vector of matrix S at row j^{th}
        distance = EuclidDistance $(T_i,S_j) = \sqrt{\sum_{d=1}^{D}\left(t_{i,d} - s_{j,d}\right)^2}$
        $w_{i,j}$ = distance + min($w_{i-1,j-1}$, $w_{i-1,j}$, $w_{i,j-1}$)
      end
    end

3.4.6 SIGN PREDICTION

Having found out the distances between the recorded sign and all reference signs, sorting them allows us to form a batch comprising the most similar signs to our record. This batch serves as a crucial indicator, helping us determine if a sign repetition occurs frequently enough to bolster our prediction's confidence. In our research, we have opted for a batch size of 2 and set the threshold to 0.8. This configuration signifies that if the same sign appears at least 2 times within the batch, we confidently generate the corresponding sign label as output. However, if the sign does not meet this frequency criterion, the output is labeled as "Unknown sign". The choice of batch size and the specific threshold values is contingent upon the availability of videos for each sign within the dataset, ensuring optimal accuracy in our predictions.

3.4.7 PHRASE GENERATION

Based upon the keywords obtained from our sign recognition model, we would pass a list of these words to form meaningful phrases or sentences with the help of spacy library in NLP and provide the sentence constructed both in form of text and audio. This phase of the model ensures that the sign input provided by the user is converted to its corresponding text and speech format. We have used the combination of Mediapipe, Dynamic Time Warping (DTW) and Natural Language Processing (NLP). The implementation of our model is described for several words and greetings in Figure 4. Also, we have described the simulation and interpretation for certain sentences in Figure 5. The simulation led to the successful interpretation from a sign language to speech, through our model. During simulation both the text and the speech form of the associated sign is obtained accurately.

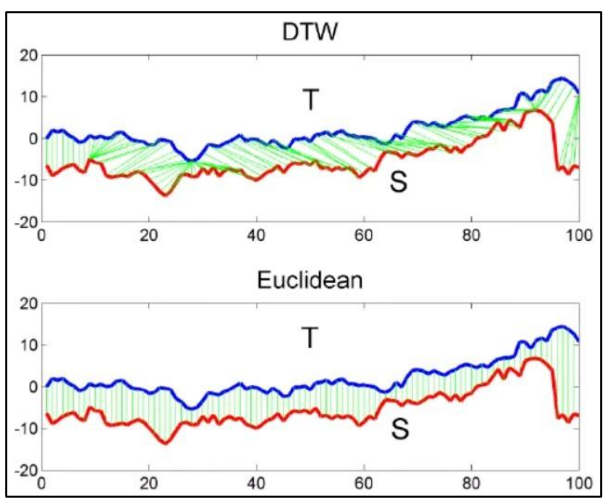

Figure-3: Comparison between DTW Distance and Euclidean Distance

CHAPTER-4

DATA COLLECTION / TOOLS /PLATFORMS USED

DATA COLLECTION/TOOLS/PLATFORMS USED

Table-2: Tools and Platforms Used

Tools / Platform	Description
USB 2.0 HD UVC Webcam	A USB 2.0 HD UVC Webcam is a high-definition video camera designed for use with computers or other devices that enable USB connectivity. Compatible with operating systems supporting USB Video Class (UVC) for plug-and-play capability; may include proprietary drivers for expanded features.
YouTube	Online video-sharing platform; allows users to publish, distribute, and view videos, including features such as subscriptions, comments, and video monetization. We used it to deploy our custom data videos on YouTube and used them as Dataset for our model
VS Code	Lightweight source code editor produced by Microsoft; supports many programming languages and provides extensions for expanded functionality.
Command Prompt	Command-line interpreter application in Windows; allows users to execute commands, explore the file system, and perform different system functions using text-based commands.
GTTS	GTTS (Google Text-to-Speech) is a Python package and API by Google; turns text into spoken words, including customizable options for voice, speed, and language. We used it to convert our obtained phrase to Speech.

CHAPTER-5

DESIGN / IMPLEMENTATION / MODELLING

DESIGN / IMPLEMENTATION / MODELLING

Our implementation of Sign Language Interpretation leverages a powerful combination of Mediapipe, Dynamic Time Warping and Natural Language Processing. The model's architecture is designed to analyze the video content in depth, with a specific focus on the pipelines created through Mediapie. This holistic approach enables our model to efficiently predict the phrases for a particular input sign, contributing to it exceptional interpretation capabilities. The workflow of implementation is divided into the following stages:

5.1 OBTAINING VIDEOS FOR REFERENCE

Our dataset mainly consists of the several Indian Sign Language gestures which are available on the website hosted by the Indian Government which comprises of multiple ISL gesture. Each and every video is on this website is hosted by YouTube. Thus, we need to extract the videos through YouTube and process them accordingly. The process of extracting the videos for training or setting us as reference video for our model is explained in Figure-4. It also explains how we would organize these videos for the further use for our model.

```
def download_video(name, video_id, start_time, duration_time):
    """
    start and end have to be in the format mm:ss
    """
    file_path = os.path.join(FOLDER, name)
    if not os.path.exists(file_path):
        os.mkdir(file_path)

    video = (
        YouTube(f"https://www.youtube.com/watch?v={video_id}")
        .streams.filter(file_extension="mp4")
        .first()
    )
    file_name = re.sub(r"[.;:,?!]", "", video.title) + ".mp4"
    if not os.path.exists(os.path.join(FOLDER, file_name)):
        video.download(FOLDER)

    output_file = os.path.join(file_path, name + "-" + video_id + ".mp4")
    if os.path.exists(output_file):
        return

    if start_time != start_time and duration_time != duration_time:
        copyfile(src=os.path.join(FOLDER, file_name), dst=output_file)
    else:
        original_video = os.path.join(FOLDER, file_name)
        try:
            os.system(
                f'ffmpeg -hide_banner -loglevel error -ss {start_time} -i "{original_video}" -to {duration_time} -c copy "{output_file}"'
            )
        except:
            print(f"An error occurred when downloading {video.title}.mp4")
            # Delete the videos used to create the clips for the dataset
            for file in os.listdir(FOLDER):
                if file.endswith(".mp4"):
                    os.remove(os.path.join(FOLDER, file))
```

Figure-4: Obtaining Reference Videos

5.2 EXTRACTING LANDMARKS

Initially, we initiate the process by extracting landmarks from the video stream utilizing the Holistic model's process method. To ensure seamless integration with OpenCV, a crucial step involves converting colors. OpenCV operates in BGR color space, whereas MediaPipe utilizes RGB colors. This color conversion is imperative for accurate rendering and analysis of the extracted landmarks. The process of landmark extraction is given in Figure-5.

```
def extract_landmarks(results):
    """Extract the results of both hands and convert them to a np array of size
    if a hand doesn't appear, return an array of zeros

    :param results: mediapipe object that contains the 3D position of all keypoints
    :return: Two np arrays of size (1, 21 * 3) = (1, nb_keypoints * nb_coordinates) corresponding to both hands
    """
    pose = landmark_to_array(results.pose_landmarks).reshape(99).tolist()

    left_hand = np.zeros(63).tolist()
    if results.left_hand_landmarks:
        left_hand = landmark_to_array(results.left_hand_landmarks).reshape(63).tolist()

    right_hand = np.zeros(63).tolist()
    if results.right_hand_landmarks:
        right_hand = (
            landmark_to_array(results.right_hand_landmarks).reshape(63).tolist()
        )
    return pose, left_hand, right_hand
```

Figure-5: Landmark Extraction

5.3 DRAWING LANDMARKS

The “drawing_utils” sub-package within MediaPipe offers a convenient toolkit encompassing all the necessary instruments needed to effortlessly visualize landmarks on a singular image. It is used to draw the connections for both left and right hand, as represented in Figure-6.

```
def draw_landmarks(image, results):
    mp_holistic = mp.solutions.holistic  # Holistic model
    mp_drawing = mp.solutions.drawing_utils  # Drawing utilities

    # Draw left hand connections
    image = mp_drawing.draw_landmarks(
        image,
        landmark_list=results.left_hand_landmarks,
        connections=mp_holistic.HAND_CONNECTIONS,
        landmark_drawing_spec=mp_drawing.DrawingSpec(
            color=(232, 254, 255), thickness=1, circle_radius=4
        ),
        connection_drawing_spec=mp_drawing.DrawingSpec(
            color=(255, 249, 161), thickness=2, circle_radius=2
        ),
    )
    # Draw right hand connections
    image = mp_drawing.draw_landmarks(
        image,
        landmark_list=results.right_hand_landmarks,
        connections=mp_holistic.HAND_CONNECTIONS,
        landmark_drawing_spec=mp_drawing.DrawingSpec(
            color=(232, 254, 255), thickness=1, circle_radius=4
        ),
        connection_drawing_spec=mp_drawing.DrawingSpec(
            color=(255, 249, 161), thickness=2, circle_radius=2
        ),
    )
    return image
```

Figure-6: Drawing the Landmarks

5.4 DTW DISTANCE CALCULATION

It is one of the most important stages of our proposed model which involves the process of finding the similarity between the reference video and the recorded video that is the live video which we are feeding to our model. Based upon the recorded video, it would return a sign directory, which is sorted on the basis of DTW distance of recorded signs to the reference videos. It is used to examine the similarity of the reference and the recorded video and return the most appropriate label for a particular sign on the basis of minimum DTW distance. The main purpose revolves around correctly predicting the sign. Figure-7 represents how the DTW Distance is to be calculated.

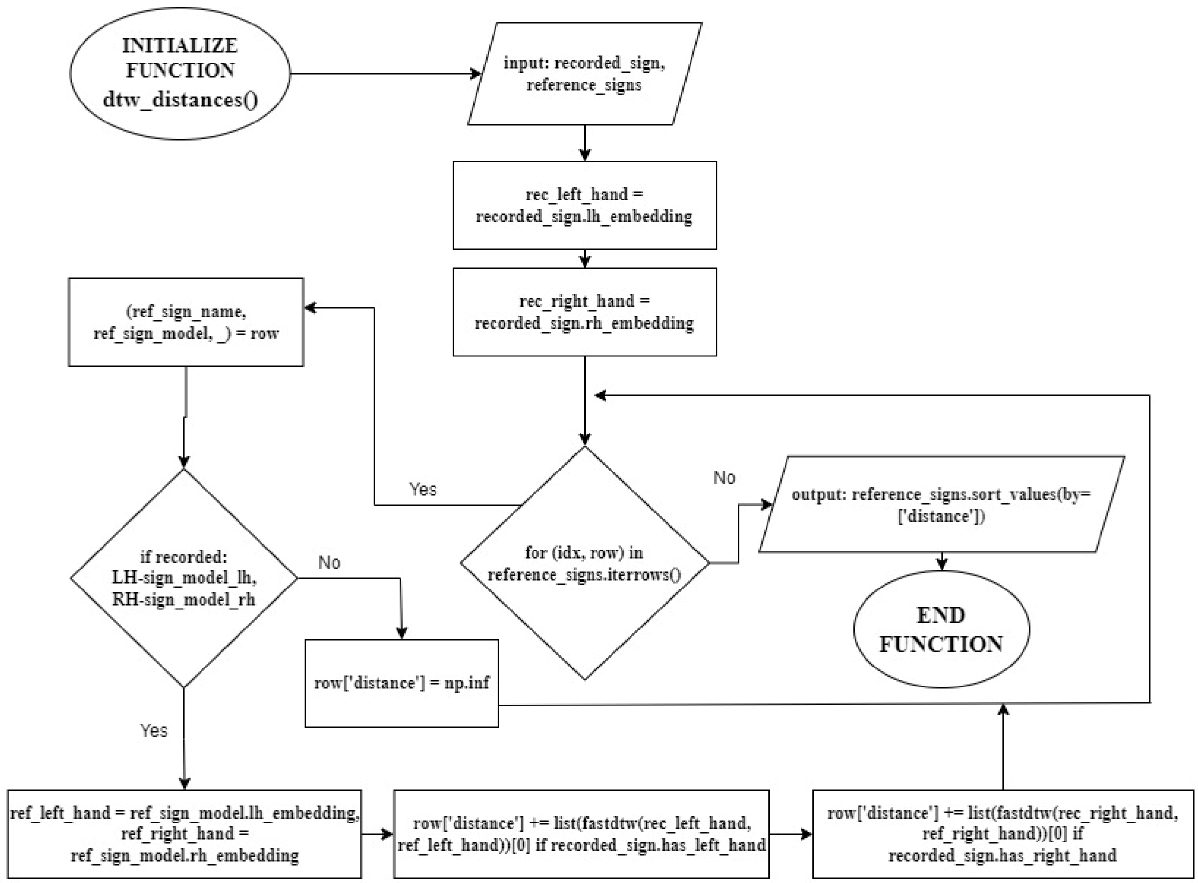

Figure-7: DTW Distance Calculation

5.5 SIGN PREDICTION

After calculating the distances between the recorded sign and all reference signs, sorting them helps us gather a batch of the most similar signs to our record. This batch allows us to verify if a sign repeats frequently enough to trust our prediction. In the following code, we've set batch_size=5 and threshold=0.5. This setup implies that if the same sign occurs at least 3 times within the batch, it becomes our output. Otherwise, we label it as "Unknown sign". These values are determined based on the quantity of videos available for each sign in the dataset, ensuring the reliability of our predictions. Implementation for sign prediction is represented in Figure-7.

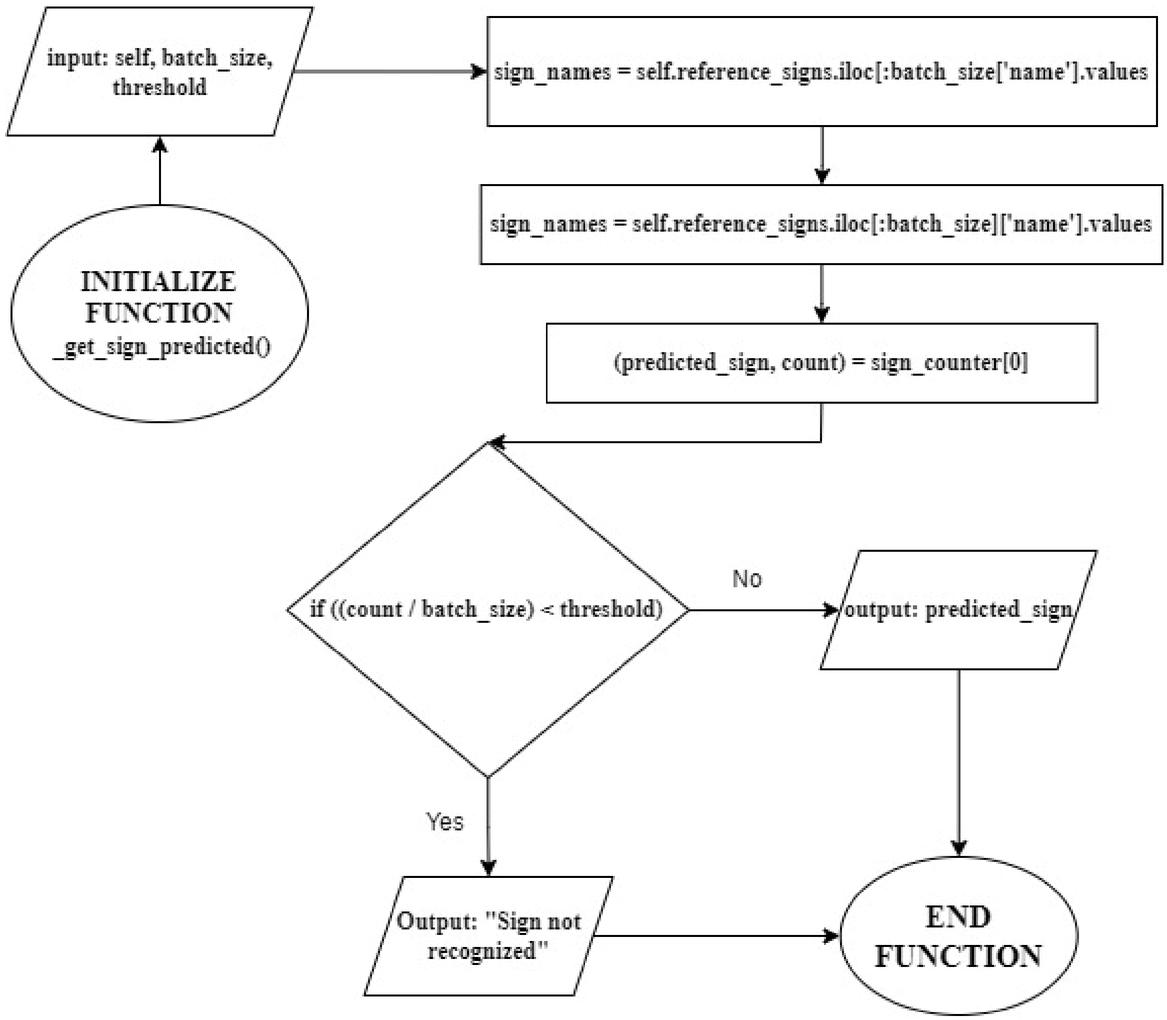

Figure-8: Sign Prediction

CHAPTER-6

TESTING AND SUMMARY OF RESULTS

TESTING AND SUMMARY OF RESULTS

Here, we will be discussing the recognition ability of our model with context to accuracy and complexity. The accuracy and complexity of both the approaches used for recognition were compared. The first method was LSTM for recognition and the other one was DTW. Model in which recognition was done through LSTM required a huge amount of data for each sign to work efficiently. To train a LSTM model, we have to run iterative epochs, and the time required to complete these epochs depends upon the size of the dataset and also on the computational power of the system. Thus, the LSTM model had a worse time and space complexity. Whereas, the model in which DTW was used for recognition, required less data as it compared similarity between two signs. Due to which model was more accurate and had less time and space complexity. On the basis of 3 Test Cases Conducted, accuracy was obtained as mentioned in Table-2.

Table-3: Model Accuracy Comparison

Test Case	**LSTM+RNN Accuracy**	**DTW Accuracy**
Test 1	80%	92%
Test 2	83%	95%
Test 3	82%	93%
Average	81.6%	93.3%

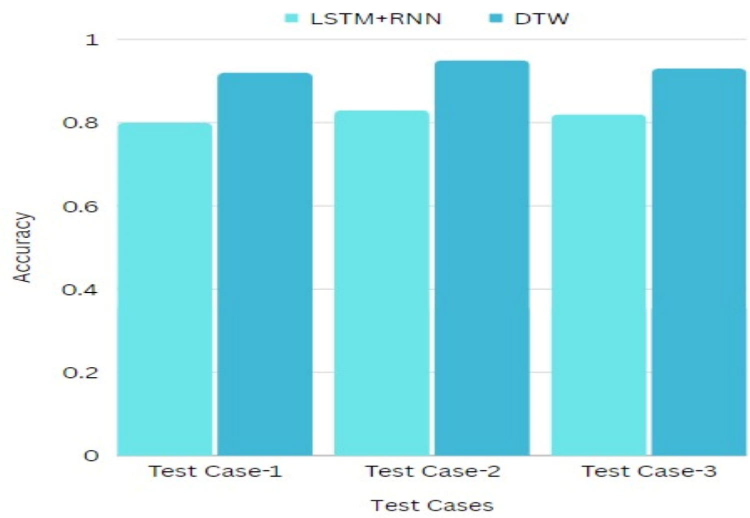

Figure-9: RNN+LSTM and DTW Accuracy Comparison

Our proposed Sign Language Interpreter was used to evaluate series of words and form sentences based on it. Our model was successfully able to identify majority of keywords and form meaningful sentences as shown in Figure-9 and Figure-10.

Figure-10: Using the combination of Mediapipe and Dynamic Time Warping to Interpretate Distinct ISL Signs

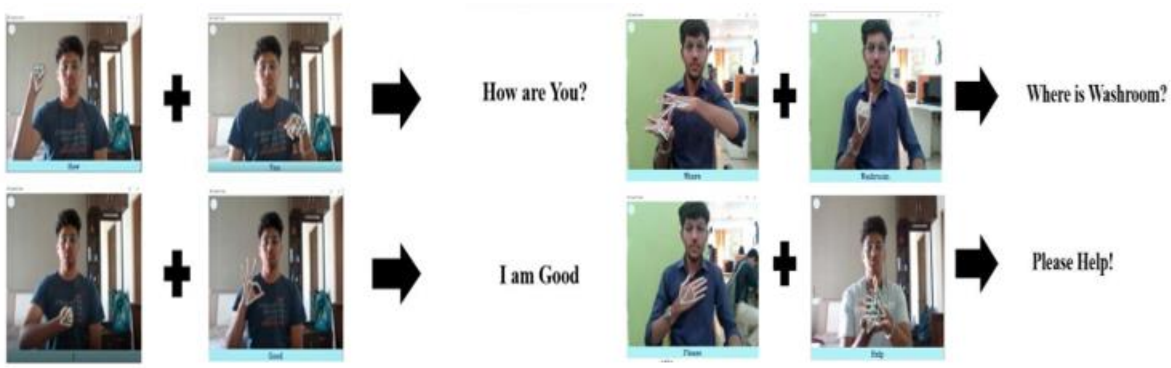

Figure-11: Using proposed Sign to Sentence Interpretation using Dynamic Time Warping to Run the Simulation against our model and generating the Sentences based on the Indian Sign Language

This study outlines a successful and comprehensive approach to sign language recognition using the MediaPipe and Dynamic Time Warping. By leveraging the model's capabilities to track hand positions, pose landmarks, and facial landmarks in real-time, the study has enabled the prediction of sign language gestures. It also provides an effective and efficient approach to sign language recognition, making it a valuable contribution to the field of computer vision and gesture recognition. The use of Dynamic Time Warping for sign comparison allows for robust recognition even under variations in sign execution speed. Additionally, the approach's reliance on feature vectors and time series analysis minimizes the need for extensive training data, making it a practical solution for real-world applications in sign language recognition.

CHAPTER-7

CONCLUSION

CONCLUSION

Based upon the model and the overall results achieved for Sign Language Interpretation, we can conclude the following:

- Successful utilization of MediaPipe and Dynamic Time Warping for sign language recognition.
- Leveraging hand positions, pose landmarks, and facial landmarks in real-time for gesture prediction.
- Contribution to computer vision and gesture recognition fields with an efficient sign language recognition approach.
- Robust recognition despite variations in sign execution speed due to Dynamic Time Warping.
- Minimized reliance on extensive training data through feature vectors and time series analysis.
- Practical applicability for real-world sign language recognition applications.
- Proposal for a model facilitating effective conversion of sign language to spoken language, enhancing communication.
- Successful conversion of input sign to its particular word or phrase in terms of text as well as audio form.

CHAPTER-8

FUTURE SCOPE

FUTURE SCOPE

- **Expansion of Sign Language Vocabulary:** The project can be extended to incorporate a bigger vocabulary of sign language signs, spanning a greater range of expressions and phrases. This would boost the system's capabilities to interpret a greater variety of sign language and communications in diverse native languages.

- **Multilingual Translation:** The technology can be improved to provide real-time translation of sign language into numerous spoken languages. By adding language models and translation algorithms, the system can accommodate varying linguistic needs, facilitating communication between sign language users and those who speak different languages.

- **Sign Recognition Refinement:** Further research and development can focus on enhancing the accuracy and robustness of the sign recognition component. By refining the algorithms and training techniques, the system can better accommodate changes in hand forms, movements, and signs, leading to more reliable sign language interpretation.

- **Integration with Mobile Devices:** The project can be extended to produce mobile applications that run the sign language recognition and interpretation system. This would enable individuals to access the system on their cellphones or tablets, enhancing convenience and accessibility in varied contexts.

- **Real-time Feedback and Correction:** Implementing a feedback mechanism within the system can allow users to examine and fix any misinterpretations or flaws in the translation. This interactive element can strengthen the system's learning capacity and further improve its accuracy over time.

CHAPTER-9

REFERENCES

REFERENCES

[1] Yifan Zhang, Ling Long, Diwei Shi, Haowen He, and Xiaoyu Liu, “Research and Improvement of Chinese Sign Language Detection Algorithm Based on YOLOv5s”, 2nd International Conference on Networking, Communications and Information Technology (NetCIT)

[2] Aakash Deep, Aashutosh Litoriya, Akshay Ingole, Vaibhav Asare, Shubham M Bhole, and Shantanu Pathak, “Realtime Sign Language Detection and Recognition”, 2nd Asian Conference on Innovation in Technology (ASIANCON)

[3] Dhruvin Gandhi, Kushal Shah and Madhav Chandane, “Dynamic Sign Language Recognition and Emotion Detection using MediaPipe and Deep Learning”, 13th International Conference on Computing Communication and Networking Technologies (ICCCNT)

[4] Gouri Anilkumar, M S Fouzia, and G S Anisha, “Imperative Methodology to Detect the Palm signs (American Sign Language) using YOLOv5 and MediaPipe”, 2nd International Conference on Intelligent Technologies (CONIT)

[5] Aishwarya D Shetty, Jyothi Shetty, Karthik K, Rakshitha, and Shabari Shedthi B, “Real-Time Translation of Sign Language for Speech Impaired”, 7th International Conference on Computing Methodologies and Communication (ICCMC)

[6] Sakshi Sharma and Sukhvinder Singh, “Recognition of Indian Sign Language (ISL) Using Deep Learning Model”, Wireless Personal Communications -An International Journal -Springer

[7] N Rajasekhar, M Geetha Yadav, Charitha Vedantam, Karthik Pellakuru, Chaitanya Navapete, “Sign Language Recognition using Machine Learning Algorithm”, International Conference on Sustainable Computing and Smart Systems (ICSCSS)

[8] B. V. Chowdary, Ajay Purshotam Thota, Alwa Sreeja, Kotla Nithin Reddy, and Karanam Sai Chandana, “Sign Language Detection and Recognition using CNN”, International Conference on Sustainable Computing and Smart Systems (ICSCSS)

[9] Shagun Katoch, Varsha Singh, and Uma Shanker Tiwary, "Indian Sign Language recognition system using SURF with SVM and CNN", International Journal of Trend in Scientific Research and Development (IJTSRD)

[10] Atharv Ganpatye and Sunil Mane, “Motion Based Indian Sign Language Recognition using Deep Learning”, 2nd International Conference on Intelligent Technologies (CONIT)

[11] Cheng, J. et al. "Chinese Sign Language Recognition Based on DTW-Distance-Mapping Features."

[12] Galván-Ruiz, J. et al. "Robust Identification System for Spanish Sign Language Based on Three-Dimensional Frame Information."

[13] Kanakanti, M. et al. "MultiFacet: A Multi-Tasking Framework for Speech-to-Sign Language Generation." International Institute of Information Technology Hyderabad, India.

[14] Kowdiki, M. and Khaparde, A. "Adaptive Hough Transform with Optimized Deep Learning Followed by Dynamic Time Warping for Hand Gesture Recognition." Springer Science+Business Media, LLC.

[15] Lu, C. et al. "Data Glove with Bending Sensor and Inertial Sensor Based on Weighted DTW Fusion for Sign Language Recognition." Graduate School of Computer Science and Engineering, University of Aizu, Japan.

[16] Hussain, M. J. et al. "Intelligent Sign Language Recognition System for E-Learning Context." Computers, Materials & Continua Tech Science Press.

[17] "Human Action Recognition Using Dynamic Time Warping and Voting Algorithm." VNU Journal of Science.

[18] Wang, J. and Zheng, H. "View-robust action recognition based on temporal self-similarities and dynamic time warping." In IEEE Int. Conf. on Computer Science and Automation Engineering (CSAE), 2012, pp. 498-502.

APPENDICES

PROJECT GUIDE:	PROF. MADHURI SAHU
GROUP MEMBERS:	MR. SHRIKANT KHAWSHE, MR. VEDANSHU THUNE, MR. YASH MANKAR, MR. YOGVID WANKHEDE
PROJECT TITLE:	SIGN LANGUAGE INTERPRETER

report

ORIGINALITY REPORT

10%	7%	4%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.sicara.fr Internet Source	2%
2	jcsce.vnu.edu.vn Internet Source	1%
3	www.researchgate.net Internet Source	1%
4	ijisrt.com Internet Source	<1%
5	Submitted to University of North Texas Student Paper	<1%
6	"Computer Vision – ECCV 2020", Springer Science and Business Media LLC, 2020 Publication	<1%
7	www.mdpi.com Internet Source	<1%
8	Submitted to CSU Northridge Student Paper	<1%
9	Aliyu, Salihu Oladimeji. "A New Approach to Arabic Sign Language Recognition System",	<1%

	King Fahd University of Petroleum and Minerals (Saudi Arabia), 2023 Publication	
10	www.coursehero.com Internet Source	<1 %
11	Hongdao Dong, Herbert Shin, Edward Ho, Helen Jingshu Jin et al. "Next-Generation Remote Hand Assessments: Cross-Platform DIGITS Web Application", Journal of Hand Surgery Global Online, 2023 Publication	<1 %
12	Submitted to Singapore Institute of Technology Student Paper	<1 %
13	Submitted to University of East London Student Paper	<1 %
14	Submitted to Asia Pacific University College of Technology and Innovation (UCTI) Student Paper	<1 %
15	dokumen.tips Internet Source	<1 %
16	ijritcc.org Internet Source	<1 %
17	Archana M R, Dayanand J, Deepak N Biradar. "Detecting Image Forgeries: A Key-Point Based Approach", 2023 International	<1 %

Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE), 2023
Publication

18	Muhammad Muddasar, Rabia Liaquat, Muhammad Zia Ur Rahman, Asif Hussain Khoja, Ayesha Aslam, Abdul Basit. "Coaxial microbial electrolysis cell for cost-effective bioenergy production and wastewater treatment of potato industry effluent", Journal of Chemical Technology & Biotechnology, 2023 Publication	<1 %
19	journals.plos.org Internet Source	<1 %
20	www.techscience.com Internet Source	<1 %
21	Chenghong Lu, Shingo Amino, Lei Jing. "Data Glove with Bending Sensor and Inertial Sensor Based on Weighted DTW Fusion for Sign Language Recognition", Electronics, 2023 Publication	<1 %
22	www.ijnrd.org Internet Source	<1 %
23	www.kluniversity.in Internet Source	<1 %

www.ujs.sk

24	Internet Source		<1 %
25	Anjana Devi V, Charulatha T, Dharishinie P. "Sign Language Recognition and Training Module", Research Square Platform LLC, 2023 Publication		<1 %
26	Giuseppe A G Lombardo, Francesco Ciancio, Hillary Violini, Claudia Sapuppo et al. "Implementing Nexobrid for Burn Debridement in a Major Italian Burn Center: A Retrospective Review of Patient Outcomes", Journal of Burn Care & Research, 2023 Publication		<1 %
27	Gouri Anilkumar, M S Fouzia, G S Anisha. "Imperative Methodology to Detect the Palm Gestures (American Sign Language) using Y010v5 and MediaPipe", 2022 2nd International Conference on Intelligent Technologies (CONIT), 2022 Publication		<1 %
28	Mayur Akewar. "Classification of EEG Signals Utilizing DWT for Feature Extraction and Evolutionary Algorithms for Feature Selection", Institute of Electrical and Electronics Engineers (IEEE), 2023 Publication		<1 %

29	Pegah Kharazmi, Jiannan Zheng, Harvey Lui, Z. Jane Wang, Tim K. Lee. "A Computer-Aided Decision Support System for Detection and Localization of Cutaneous Vasculature in Dermoscopy Images Via Deep Feature Learning", Journal of Medical Systems, 2018 Publication	<1 %
30	Yifan Zhang, Ling Long, Diwei Shi, Haowen He, Xiaoyu Liu. "Research and Improvement of Chinese Sign Language Detection Algorithm Based on YOLOv5s", 2022 2nd International Conference on Networking, Communications and Information Technology (NetCIT), 2022 Publication	<1 %
31	Zheng Yu Tan, Shafriza Nisha Basah, Haniza Yazid, Muhammad Juhairi Aziz Safar. "Performance analysis of Otsu thresholding for sign language segmentation", Multimedia Tools and Applications, 2021 Publication	<1 %
32	dokumen.pub Internet Source	<1 %
33	link.springer.com Internet Source	<1 %
34	pure.mpg.de Internet Source	<1 %

35	www.frontiersin.org Internet Source	<1 %
36	www.itopvpn.com Internet Source	<1 %
37	www.ncbi.nlm.nih.gov Internet Source	<1 %
38	Madise, Kabelo Germanus. "A Multi Sensor Approach to Botswana Sign Language Dataset With View of Addressing Occlusion", Botswana International University of Science & Technology (Botswana), 2021 Publication	<1 %
39	"Advances in Computing and Data Sciences", Springer Science and Business Media LLC, 2021 Publication	<1 %
40	Hafiz Muhammad Hamza, Aamir Wali. "Pakistan sign language recognition: leveraging deep learning models with limited dataset", Machine Vision and Applications, 2023 Publication	<1 %

Exclude quotes On
Exclude bibliography Off
Exclude matches Off

Real-Time Sign Language Interpreter using Mediapipe, Dynamic Time Warping and NLP

Yash Mankar[1], Vedanshu Thune[2], Shrikant Khawshe[3], Yogvid Wankhede[4] and Madhuri Sahu[5]
[1] Department of Artificial Intelligence, G.H. Raisoni College of Engineering, Nagpur, India
Email: yashmankar1120@gmail.com
[2-5] Department of Artificial Intelligence, G.H. Raisoni College of Engineering, Nagpur, India
Email: {vedanshuthune, shrikantkhawshe727, yogvidwankhede, madhuridec9}@gmail.com

Abstract. The "Sign Language Interpreter" project is dedicated to alleviating communication challenges encountered by individuals who're unable to express themselves through means. This innovative study utilizes hand gestures to generate sentences in time thereby facilitating effective communication. Inspired by the language of signs our system employs algorithms for gesture recognition. It. Translates the nuances of hand movements into sentences ensuring accuracy and contextual understanding. The Sign Language Interpreter incorporates cutting edge technologies to comprehend sign language intricacies.

Index Terms. Indian Sign-Language, Hand Gesture Recognition, Mediapipe, Dynamic Time Warping, Landmark Extraction, NLP

I. Introduction

The proposed study urges to implement a model that detects the hand gesture and in order to interpret those conversations to any existing written and spoken languages. A real-time sign language interpreter using Mediapipe and Dynamic Time Warping is being developed. Sign language is a highly expressive and intricate form of communication that combines fluid hand movements, facial expressions, head movements, and body language. It is the seamless integration of these elements that makes it a complete and effective means of communication. However, this presents a formidable task in the field of computer vision, primarily due to the complexity of data extraction and subsequent analysis. Furthermore, each individual possesses their unique body shape, volume, and distinctive gesture style, thereby complicating the recognition process. Moreover, the presence of uncertainties such as variations in viewpoint, illumination, shadows, self-occlusion, deformation, noise, and clothing further intensify the complexity of this problem.

Our primary goal is to provide an accessible solution for individuals with speech impairments enabling effortless communication with an audience. In addition to converting sign language into spoken words, the proposed sign language interpreter offers features such as an affordable and portable user interface making it suitable for sign recognition and converting those to spoken words as well as phrases. The interpreter is designed to be user friendly so that individuals with levels of proficiency can communicate seamlessly using sign language. This research paper outlines the development process, underlying technologies used in creating the Sign Language Interpreter, as its evaluation. Our goal for this project is to make a contribution to technology promoting better communication and inclusiveness, for people who have difficulty speaking by combining technical advancements such as Mediapipe, Dynamic Time Warping and Natural Language Processing.

At the center of our action lies the application of a comprehensive Indian Sign Language database as well as our custom dataset for the training of our model. This expansive database serves as the bedrock for developing our system. The addition of Indian Sign Language ensures that our model is culturally sensitive, understanding the unique verbal nuances and expressions. The variance in the performance of human actions over time results in unique feature representations for each action in different samples. This poses a challenge for widely-used classifiers like neural networks and support vector machines, which typically expect fixed-size input feature vectors and struggle with action recognition tasks. The research introduces a classification model using Dynamic Time Warping (DTW) and a voting algorithm to address the challenge of unique feature representations in human actions. The system also uses natural language processing to convert text to speech for spoken language, enhancing its effectiveness and inclusivity. This groundbreaking design in assistive technology is a groundbreaking effort.

II. RELATED WORK

These existing systems aim to enable communication among individuals who generally use sign languages and common people. One of the main challenges in developing sign language recognition systems is the lack of sufficient data for training machine learning models. To address this issue, various models have been proposed for collecting and recycling sign language data, such as using depth cameras, marker- grounded systems, and wearable detectors. Another challenge is the higher variance in sign language gestures, which makes it delicate to train models to recognize specific signs with high efficiency. In terms of speech conversion, some studies have used Text- to- Speech (TTS) has been employed for speech conversion, however it has limitations such as a lack of persuasive voice and restricted language support. More research is needed to develop reliable sign recognition models and improve voice conversion quality.

In Sign Language Recognition, there have been numerous former exploration papers that have proposed different styles and ways to improve the delicacy of the recognition system. A comparison of crucial features from former exploration and a specific paper in Sign Language Recognition would involve looking at factors similar as the dataset used, the AI model armature, the input and affair, the evaluation criteria, and the results achieved. For illustration, a comparison of crucial features between former exploration and a specific paper in Sign Language Recognition could include:

i. Dataset for prior exploration may have used a different dataset than the specific paper, which can affect the results achieved.

ii. Input and Output for former exploration may have used different input and affair formats than the specific paper, which can affect the performance of the recognition system.

iii. Evaluation Criteria in prior exploration may have used different evaluation criteria than the specific paper, which can affect the interpretation of the results.

iv. Results achieved by former exploration may have achieved different results than the specific paper, which can affect the overall conclusion of the exploration.

In conclusion, a comparison of crucial features between former exploration and a specific paper in subscribe Language Recognition would involve looking at factors similar as the dataset used, the input and affair, the evaluation criteria, and the results achieved. Understanding these parallels and differences.

III. PROPOSED METHODOLOGY

This section explains the proposed methodology in detail. The proposed model of a sign language recognition to speech interpretation using our methodology involves several key steps, culminating in the implementation of a real-time sign language interpreter utilizing Dynamic Time Warping and Mediapipe technologies. Illustrated in Figure-1, our Dynamic Time Warping Classification model serves as the cornerstone of the system, aimed at generating corresponding words based on sign language recorded in a video recorded through the camera. Our Sign Language Recognition model mainly focuses on the hand landmarks, representing significant points of interest on the hand, which we meticulously track and utilize as input data. The prediction process in our model unfolds in the following manner:

i. Extract Landmarks
ii. Calculate the DTW distance between the recorded and reference signs.
iii. Predict the sign/gesture by analyzing which reference sign is likely to be recorded.

A. Problem domain

The domain encompasses the development of a comprehensive system capable of real-time detection and interpretation of hand gestures from sign language, bridging the communication gap between individuals with speech impairments and those without.

B. Problem Defination

Creating an effective sign language interpreter using Mediapipe and Dynamic Time Warping, capable of instantaneously translating sign language into spoken and written languages. This involves overcoming the complexities inherent in sign language recognition due to the non-rigid nature of the human body, individual variations in gestures, viewpoint uncertainties, and environmental factors like illumination and occlusion.

C. Problem Statement

To create an accessible solution for individuals with speech impairments, facilitating seamless communication with diverse audiences. Beyond converting sign language into spoken words, the proposed interpreter aims to offer affordability, portability, and user-friendliness, accommodating varying levels of proficiency in sign language users. By integrating Mediapipe, Dynamic Time Warping, and Natural Language Processing (NLP), this research endeavors to contribute to inclusive technology, addressing communication barriers faced by those with speech disabilities.

D. Dataset

The Indian government hosts an extensive Indian Sign Language (ISL) dataset on its website, encompassing a rich repository of sign language gestures. This collection comprises diverse gestures, expressions, and linguistic nuances vital for comprehensive sign language interpretation. Developed to ensure cultural sensitivity, this dataset serves as a fundamental resource for training models aimed at enhancing communication accessibility for individuals reliant on ISL as their primary mode of expression. We have also included our own sign language videos to train our model which include the various words as represented in TABLE I.

TABLE I: CUSTOM TRAINING DATASET

Index	Action Type	Training Sample set
1	How	5
2	You	5
3	I	5
4	Good	5
5	What	5
6	Time	5
7	Where	5
8	Washroom	5
9	Please	5
10	Help	5

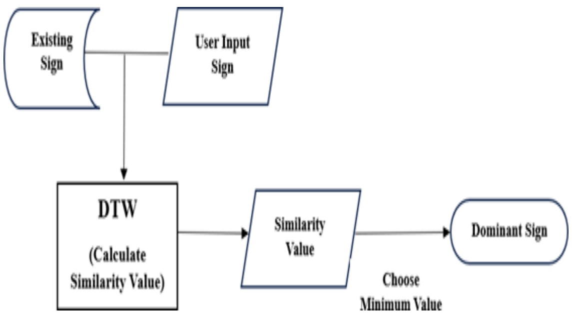

Figure 1. Dynamic Time Warping Classifier

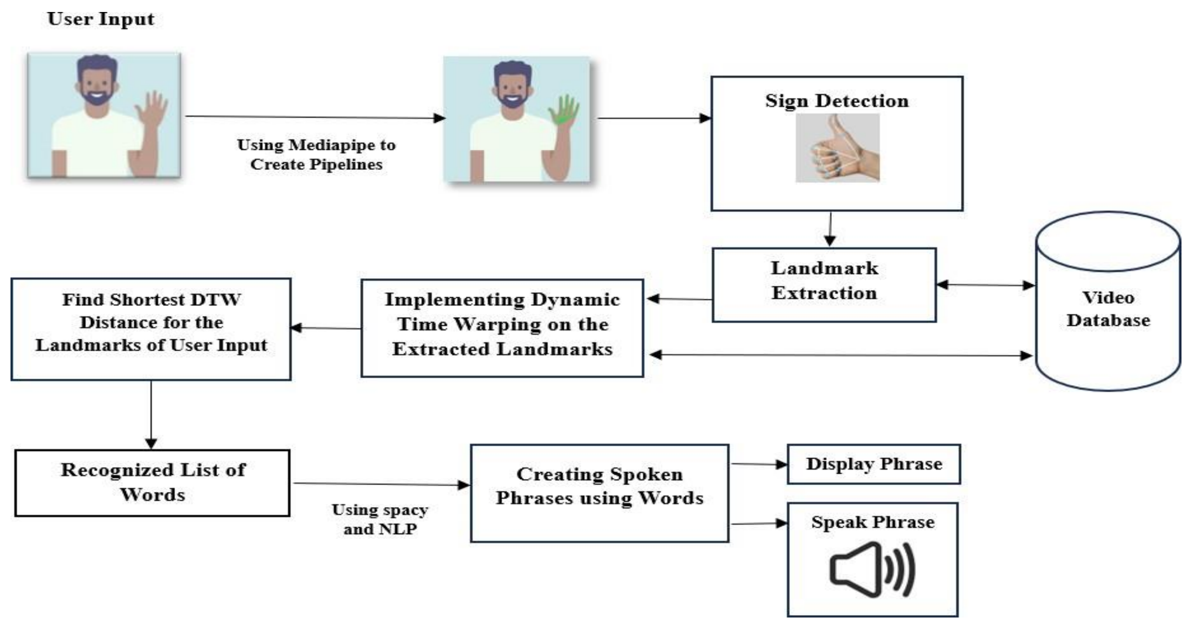

Figure 2. Proposed Model for Sign language Interpretaion using Dynamic Time Warping

IV. IMPLEMENTATION

The proposed Dynamic Time Warping (DTW) classifies and contrast time series data effectively. The focus lies in calculating the similarity between two time series. Its adaptability for accommodating time series variances positions it as an indispensable asset within the realms of machine learning and data analysis. Here, we are comparing the existing sign and the user input sign for classification. For this study, we are considering our own custom dataset for training our model. The contents of our model are given in TABLE I. These are some of the basic sign language gestures which are widely used for communication. It also mentions the number of training samples for them respectively. Also the implemented system for sign language interpretation is given in Figure.2.

A. MediaPipe Detection/Holistic Model

MediaPipe is an open-source computer vision framework developed by Google a few years ago. Within the suite of models offered by MediaPipe, the Holistic Model stands out, enabling real-time tracking of Hand, Pose, and Face landmarks. In our current model, we leverage Hand positions and Pose landmarks, which include the shoulder, elbow, and wrist positions, to inform our predictions.

B. Extract landmarks

Initially, we initiate the process by extracting landmarks from the video stream utilizing the Holistic model's process method. To ensure seamless integration with OpenCV, a crucial step involves converting colors. OpenCV operates in BGR color space, whereas MediaPipe utilizes RGB colors. This color conversion is imperative for accurate rendering and analysis of the extracted landmarks.

C. Draw landmarks

The “drawing_utils” sub-package within MediaPipe offers a convenient toolkit encompassing all the necessary instruments needed to effortlessly visualize landmarks on a singular image.

D. Models

The project's current interpretation involves the execution of two distinct models. The "HandModel" class encapsulates detailed information about hand gestures within a single image. Meanwhile, the "SignModel" class encompasses and extends the "HandModel" data that is seamlessly applied across all frames within a video, ensuring comprehensive coverage and accurate analysis throughout the entire footage.

E. *Dynamic Time Warping (DTW)*

It is a widely employed algorithm specifically designed for comparing time series data. It excels in identifying optimal alignments between two time series through a warping technique, allowing for the analysis of patterns rather than mere sequences. In our context, DTW proves invaluable in recognizing the similarities between embeddings of the same signs. This method enables us to identify the most similar signs based on their DTW distances, providing a reliable means of sign recognition. The resulting approach furnishes a comprehensive list of signs, sorted in order of their proximity to the recorded sign. This sorting is achieved by their respective distances calculated using the Dynamic Time Warping technique. For detailed understanding of the DTW calculation process, we need to consider 2 matrixes to find similarity among them. The following represents how to use Dynamic Time Warping to find out the similarity score:

Input: **T** = $[t_{i,d}]_{MxD}$ and **S**= $[s_{i,d}]_{MxD}$
Output: similarity between two input action matrices
function matrixsimlilarity(T,S)
let $[w_{i,j}]_{MxN}$ be warping matrix aligning two feature representations
set $w_{i,j}$ infinity for all i and j
for $i=1$ to M do
for $j=1$ to N do
let T_i be the row vector of matrix T at row i^{th}
let S_j be the row vector of matrix S at row j^{th}
distance = EuclidDistance $(T_i,S_j) = \sqrt{\sum_{d=1}^{D}\left(t_{i,d} - s_{j,d}\right)^2}$
$w_{i,j}$ = distance + min($w_{i-1,j-1}$, $w_{i-1,j}$, $w_{i,j-1}$)
end
end
return $w_{M,N}$
end function

F. *Sign Prediction*

Having found out the distances between the recorded sign and all reference signs, sorting them allows us to form a batch comprising the most similar signs to our record. This batch serves as a crucial indicator, helping us determine if a sign repetition occurs frequently enough to bolster our prediction's confidence. In our research, we have opted for a batch size of 2 and set the threshold to 0.8. This configuration signifies that if the same sign appears at least 2 times within the batch, we confidently generate the corresponding sign label as output. However, if the sign does not meet this frequency criterion, the output is labeled as "Unknown sign". The choice of batch size and the specific threshold values is contingent upon the availability of videos for each sign within the dataset, ensuring optimal accuracy in our predictions.

G. *Phrase Generation*

Based upon the keywords obtained from our sign recognition model, we would pass a list of these words to form meaningful phrases or sentences with the help of spacy library in NLP and provide the sentence constructed both in form of text and audio. This phase of the model ensures that the sign input provided by the user is converted to its corresponding text and speech format. We have used the combination of Mediapipe, Dynamic Time Warping (DTW) and Natural Language Processing (NLP). The implementation of our model is described for several words and greetings in Figure 4. Also, we have described the simulation and interpretation for certain sentences in Figure 5. The simulation led to the successful interpretation from a sign language to speech, through our model. During simulation both the text and the speech form of the assoiciated sign is obtained accurately.

V. RESULT

Here, we will be discussing the recognition ability of our model with context to accuracy and complexity. The accuracy and complexity of both the approaches used for recognition were compared. The first method was LSTM for recognition and the other one was DTW. Model in which recognition was done through LSTM required a huge amount of data for each sign to work efficiently. To train a LSTM model, we have to run iterative epochs, and the time required to complete these epochs depends upon the size of the dataset and also on the computational power of the system. Thus, the LSTM model had a worse time and space complexity. Whereas, the model in which DTW was used for recognition, required less data as it compared similarity between two signs. Due to which model was more accurate and had less time and space complexity. On the basis of 3 Test Cases Conducted, accuracy was obtained as mentioned in TABLE-2.

Figure 4. Using the combination of Mediapipe and Dynamic Time Warping to Interpretate Distinct ISL Signs

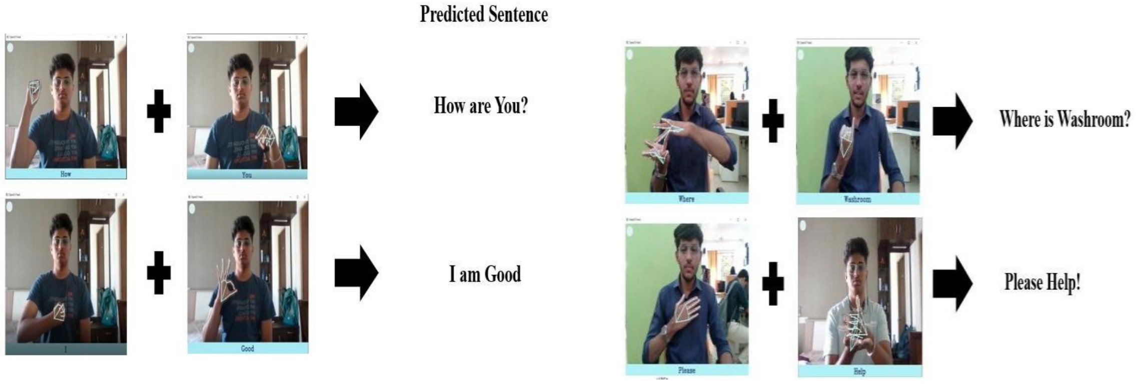

Figure 5(a)

Figure 5(b)

Figure 5. Using proposed Sign to Sentence Interpretation using Dynamic Time Warping to Run the Simulation against our model and generating the Sentences based on the Indian Sign Language

TABLE-2: MODEL ACCURACY COMPARISON

Test Case	LSTM+RNN Accuracy	DTW Accuracy
Test 1	80%	92%
Test 2	83%	95%
Test 3	82%	93%
Average	81.6%	93.3%

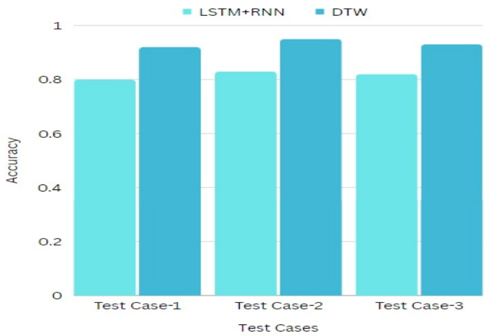

Figure 6. Comparison of Accuracy of LSTM+RNN and DTW

VI. DISCUSSION

We have developed a state-of-art sign language to speech interpreter using Dynamic Time Warping, Mediapie and NLP for Indian Sign Language (ISL). It can be used for the conversion of given sign to its respective conversion in spoken language. The present sign language generation model has shown promising results based on DTW scores. However, evaluating its real-world performance requires input from expert sign language interpreters. Assessing coherence and synchronization of non-manual elements like body posture, head movements, and eye gaze is pivotal. While the model adeptly learns broad hand movements, it lacks the finesse in finger and facial movements. One potential remedy involves exploring alternative loss functions such as keypoint loss. Also, the limitations in dataset size and diversity might constrain the model's ability to capture the intricate nuances of sign language. Investigating how different signer styles affect the model's output and devising adaptive strategies for varying signing styles is imperative for practical implementation.

VII. CONCLUSION

This research paper outlines a successful and comprehensive approach to sign language recognition using the MediaPipe and Dynamic Time Warping. By leveraging the model's capabilities to track hand positions, pose landmarks, and facial landmarks in real-time, the study has enabled the prediction of sign language gestures. It also provides an effective and efficient approach to sign language recognition, making it a valuable contribution to the field of computer vision and gesture recognition. The use of Dynamic Time Warping for sign comparison allows for robust recognition even under variations in sign execution speed. Additionally, the approach's reliance on feature vectors and time series analysis minimizes the need for extensive training data, making it a practical solution for real-world applications in sign language recognition. The study proposes a model that would successfully recognize and convert the input sign language to a spoken language enabling effective communication.

VIII. REFERENCES

[1] Yifan Zhang, Ling Long, Diwei Shi, Haowen He, and Xiaoyu Liu, "Research and Improvement of Chinese Sign Language Detection Algorithm Based on YOLOv5s", 2nd International Conference on Networking, Communications and Information Technology (NetCIT)

[2] Aakash Deep, Aashutosh Litoriya, Akshay Ingole, Vaibhav Asare, Shubham M Bhole, and Shantanu Pathak, "Realtime Sign Language Detection and Recognition", 2nd Asian Conference on Innovation in Technology (ASIANCON)

[3] Dhruvin Gandhi, Kushal Shah and Madhav Chandane, "Dynamic Sign Language Recognition and Emotion Detection using MediaPipe and Deep Learning", 13th International Conference on Computing Communication and Networking Technologies (ICCCNT)

[4] Gouri Anilkumar, M S Fouzia, and G S Anisha, "Imperative Methodology to Detect the Palm signs (American Sign Language) using YOLOv5 and MediaPipe", 2nd International Conference on Intelligent Technologies (CONIT)

[5] Aishwarya D Shetty, Jyothi Shetty, Karthik K, Rakshitha, and Shabari Shedthi B, "Real-Time Translation of Sign Language for Speech Impaired", 7th International Conference on Computing Methodologies and Communication (ICCMC)

[6] Sakshi Sharma and Sukhvinder Singh, "Recognition of Indian Sign Language (ISL) Using Deep Learning Model", Wireless Personal Communications -An International Journal -Springer

[7] N Rajasekhar, M Geetha Yadav, Charitha Vedantam, Karthik Pellakuru, Chaitanya Navapete, "Sign Language Recognition using Machine Learning Algorithm", International Conference on Sustainable Computing and Smart Systems (ICSCSS)
[8] B. V. Chowdary, Ajay Purshotam Thota, Alwa Sreeja, Kotla Nithin Reddy, and Karanam Sai Chandana, "Sign Language Detection and Recognition using CNN", International Conference on Sustainable Computing and Smart Systems (ICSCSS)
[9] Shagun Katoch, Varsha Singh, and Uma Shanker Tiwary, "Indian Sign Language recognition system using SURF with SVM and CNN", International Journal of Trend in Scientific Research and Development (IJTSRD)
[10] Atharv Ganpatye and Sunil Mane, "Motion Based Indian Sign Language Recognition using Deep Learning", 2nd International Conference on Intelligent Technologies (CONIT)
[11] Kotkondawar, R.R., Sutar, S.R., Kiwelekar, A.W., Wankhede, H.S., "Performability of Deep Recurrent Neural Networks for Molecular Sequence data", International Journal of Computing and Digital Systems, 2023, 13(1), pp. 1317–1327
[12] Laddha, M.D., Netak, L.D., Wankhede, H.S., "A Correlation Analysis Between Cognitive Process and Knowledge Dimension in Software Engineering by Using the Revised Bloom's Taxonomy", Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2021, 12615 LNCS, pp. 34–41
[13] Sahu, M., Dhawale, K., Bhagat, D., Wankhede, C., Gajbhiye, D., "Convex Hull Algorithm based Virtual Mouse", 14th International Conference on Advances in Computing, Control, and Telecommunication Technologies, ACT 2023, 2023, 2023-June, pp. 846–851
[14] Sahu, M., Francis, N.A., Sahu, A., Ulhe, P., "Data Wrangling and Visualization for Credibility and Usability", Proceedings - 2022 6th International Conference on Intelligent Computing and Control Systems, ICICCS 2022, 2022, pp. 1834–1840
[15] Nirale, P., Madankar, M., "Design of an IoT Based Ensemble Machine Learning Model for Fruit Classification and Quality Detection", International Conference on Emerging Trends in Engineering and Technology, ICETET, 2022, 2022-April
[16] Madankar, M., Chandak, M., Chavhan, N., "Information Retrieval System Based on Query Translation Approach for Cross-Languages", Lecture Notes in Electrical Engineering, 2021, 700, pp. 1261–1269
[17] Sonekar, S.V., Pal, M., Tote, M., Sawwashere, S., Zunke, S., "Computation termination and malicious node detection using finite state machine in mobile adhoc networks", Proceedings of the 7th International Conference on Computing for Sustainable Global Development, INDIACom 2020, 2020, pp. 156–161, 9083710
[18] Pal, M.M., Jasutkar, R.W., "Implementation of hand vein structure authentication-based system", Proceedings - International Conference on Communication Systems and Network Technologies, CSNT 2012, 2012, pp. 114–118, 6200601
[19] Agarwal, M., Kotecha, A., Deolalikar, A., ...Yadav, R.K., Thomas, A., "Deep Learning Approaches for Plant Disease Detection: A Comparative Review", 2023 IEEE International Students' Conference on Electrical, Electronics and Computer Science, SCEECS 2023, 2023
[20] Agarwal, M., Kalia, R., Bahel, V., Thomas, A., "AutoEval: A NLP Approach for Automatic Test Evaluation System", 2021 IEEE 4th International Conference on Computing, Power and Communication Technologies, GUCON 2021, 2021

PAPER

ORIGINALITY REPORT

5%	2%	3%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Liverpool John Moores University Student Paper	2%
2	Mounika Kanakanti, Shantanu Singh, Manish Shrivastava. "MultiFacet: A Multi-Tasking Framework for Speech-to-Sign Language Generation", International Cconference on Multimodal Interaction, 2023 Publication	1%
3	pubmed.ncbi.nlm.nih.gov Internet Source	1%
4	eprints.covenantuniversity.edu.ng Internet Source	<1%
5	Archana M R, Dayanand J, Deepak N Biradar. "Detecting Image Forgeries: A Key-Point Based Approach", 2023 International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE), 2023 Publication	<1%

ijece.iaescore.com

6	Internet Source	<1 %
7	pt.scribd.com Internet Source	<1 %

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	Off		

RAISONI GROUP
a vision beyond

Yash Mankar <yash.mankar.ai@ghrce.raisoni.net>

ICMSLE 2024- Acceptance Letter

1 message

icmsle conf <conference.icmsle@gmail.com> Sat, Nov 25, 2023 at 11:56 AM
To: Yash Mankar <yashmankar1120@gmail.com>, yash.mankar.ai@ghrce.raisoni.net, vedanshu.thune.ai@ghrce.raisoni.net, shrikant.khawshe.ai@ghrce.raisoni.net, yogvid.wankhede.ai@ghrce.raisoni.net, madhuri.sahu@raisoni.net

Dear Author

Greetings from ICMSLE 2024!

Your paper was accepted and recommended for inclusion in the 2024 Springer International Conference on Multi-Strategy Learning Environment ICMSLE-2024.

Herewith, the conference committee of the International Conference on Multi-Strategy Learning Environment **[ICMSLE-2024]** is pleased to inform you that the peer reviewed research paper has been accepted for oral presentation. **ICMSLE-2024** will be held on 12-13, January 2024 at Dehradun, India. **ICMSLE** encourages only the active participation of highly qualified delegates to bring you various innovative research ideas.

Details of Accepted Paper

Paper ID	Paper Title	Author(s)
ICMSLE011	Real-Time Sign Language Interpreter using Mediapipe, Dynamic Time Warping and NLP	Yash Mankar, Vedanshu Thune, Shrikant Khawshe, Yogvid Wankhede, Madhuri Sahu

We congratulate you on being successfully selected for the presentation of your research work in our esteemed conference.

Please find the detailed reviewer comments at the end of this email.
Last date for registration: 2 December 2023

Conference Date: 12-13, January 2024

Important Guidelines for ICMSLE 2024
Please ensure the following before uploading the final paper.

- ✓ Ensure all the figures and table to be cited in the text.
- ✓ Do not cite any references in the Abstract and Conclusion Section
- ✓ Minimum 12 references must be in the paper and all references must be cited in the text. Like [1], [2] ...
- ✓ The article has few typographical errors which may be carefully looked at.
- ✓ Spell out each acronym the first time used in the body of the paper.

FORM XIV
APPLICATION FOR REGISTERATION OF COPYRIGHT
[SEE RULE 70]

Diary Number: 29151/2023-CO/L

To

The Registrar of Copyrights,
Copyright Office,
Department of Industrial Policy & Promotion,
Ministry of Commerce and Industry,
Boudhik Sampada Bhawan,
Plot No. 32, Sector 14, Dwarka,
New Delhi-110075
Email Address: copyright@nic.in
Telephone No.: (Office) 011-28032496, 08929474194

Sir,

In Accordance with Section 45 of the Copright Act, 1957 (14 of 1957), I hereby apply for registration of Copyright and request that enteries may be made in the Register of Copyrights as in the enclosed Statement of Particulars.

1. I also send herewith duly completed the Statement of further Particulars relating to the work. (for Literary/Dramatic, Musical, Atristic works only) **Literary/ Dramatic works**

2. In accordance with rule 16 of the Copyright Rules, 1958, I have sent by prepaid registered post copies of this letter and of the Statement of Particulars and Statement of Further Particulars to other parties concerned as shown below:

Name of Party	Address of Party	Date of Dispatch
VEDANSHU THUNE	COE AIML, AI DEPARTMENT GH RAISONI COLLEGE OF ENGINEERING NAGPUR-440016	03/11/2023
YASH MANKAR	COE AIML, AI DEPARTMENT GH RAISONI COLLEGE OF ENGINEERING NAGPUR-440016-440016	03/11/2023
YOGVID WANKHEDE	COE AIML, AI DEPARTMENT GH RAISONI COLLEGE OF ENGINEERING NAGPUR-440016-440016	03/11/2023
SHRIKANT KHAWSHE	COE AIML, AI DEPARTMENT GH RAISONI COLLEGE OF ENGINEERING NAGPUR-440016-440016	03/11/2023
MADHURI SAHU	COE AIML, AI DEPARTMENT GH RAISONI COLLEGE OF ENGINEERING NAGPUR-440016-440016	03/11/2023

[See columns 7,11,12, and 13 of the Statement of Particulars and party referred in col.2 (e) of the Statement of Further Particulars.]

3. The prescribed fee has been paid, as per details below: **500/-**

Payment ID	Payment Date	Amount	Bank Name	Payment Mode
328019	03/11/2023	500		

4. Communications on this subject may be addressed to:

VEDANSHU THUNE
COE AIML, AI DEPARTMENT GH RAISONI COLLEGE OF ENGINEERING NAGPUR-440016-440016
7083088795

5. I hereby declare that to the best of my knowledge and belief, no person, other than to whom a notice has been sent as per paragraph 2 above any claim or interest or dispute to my copyright of this work or its use by me.

6. I hereby verify that the particulars given in this Form and the Statement of Particulars and Statment of Further Particulars are true to the best of my knowledge, belief and information and nothing has been concealed there from.

List of Enclosures:

1. 2 Copies of Work

2. DD/IPO of Rs.500 Per Work
3. Authorization from author/publisher
4. If the application is being filed through attorney , a specific Power of Attorney in original duly signed by the applicant and accepted by the attorney

Place:

Date: **03/11/2023**

For : VEDANSHU THUNE

Proprietor

STATEMENT OF PARTICULARS

Diary Number: 29151/2023-CO/L

1.	Registration Number	
2.	Name, Address and Nationality of the Applicant	NAME: VEDANSHU THUNE, ADDRESS: COE AIML, AI DEPARTMENT GH RAISONI COLLEGE OF ENGINEERING NAGPUR-440016-440016, Indian
3.	Nature of the Applicant's interest in the Copyright of the work	Author
4.	Class and description of the work	Literary/ Dramatic Work
5.	Title of the work	Real Time Sign Language To Speech Interpreter using Dynamic Time Warping, Mediapipe and Natural Language Processing
6.	Language of the work	English
7.	Name, Address and Nationality of the Author and if the Author is deceased, the date of decease.	NAME: VEDANSHU THUNE, ADDRESS: COE AIML, AI DEPARTMENT GH RAISONI COLLEGE OF ENGINEERING NAGPUR-440016-440016, Indian,
8.	Whether the work is Published or Unpublished	Unpublished
9.	Year and Country of first publication, and Name, Address and Nationality of the publisher	N/A
10.	Year and Countries of subsequent publications, if any, and Name, Address and Nationality of the publisher	N/A
11.	Name, Address and Nationality of the Owners of the various rights comprising the copyright in the work and extent of rights held by each, together with particulars of assignments and licence. If any	NAME: VEDANSHU THUNE, ADDRESS: COE AIML, AI DEPARTMENT GH RAISONI COLLEGE OF ENGINEERING NAGPUR-440016-440016, Indian
12.	Name and address and nationality of other persons, if any authorized to assign or licence the rights comprising the copyright	NAME: VEDANSHU THUNE, ADDRESS: COE AIML, AI DEPARTMENT GH RAISONI COLLEGE OF ENGINEERING NAGPUR-440016-440016, Indian
13.	If the work is an 'Artistic work', the location of the original work, including name, address and nationality of the person in possession of the work, (In the case of an architectural work, the year of completion of the work should also be shown)	N/A
14.	If the work is an 'Artistic work' which is used or capable of being used in relation to any goods or services, the application should include a certification from the Registrar of Trade Marks in terms of the provision to Sub-Section (i) of Section 45 of the Copyright Act, 1957	N/A
15.	If the work is an 'Artistic work' whether it is registered under the Desings Act 2000 if yes give details.	N/A
16.	If the work is an 'Artistic work' capable of being registrar as a design under the Designs Act 2000, whether is has been applied to an article though an industrial process and,if yes ,then number of times it is reproduced	N/A
17.	Remarks, if any	

Place:
Date: **03/11/2023**

For : VEDANSHU THUNE

Proprietor

STATEMENT OF FURTHER PARTICULARS

(For Literary/Dramatic, Musical and Artistic works only)

Diary Number: 29151/2023-CO/L

1. Is the work to be registered

(a) an orginal work?	: Yes
(b) a translation of a work in the public domain?	: N.A.
(c) a translation of a work in which Copyright subsists?	: N.A.
(d) an adaptation of a work in the public domain?	: N.A.
(e) an adaptation of a work in which Copyright subsists?	: N.A.

2. If the work is a translation or adaptation of a work in which copyright subsists

(a) Title of the original work	: N.A.
(b) Language of the original work	: N.A.
(c) Name, address, and nationality of the author of the original work and if the author is deceased, the date of decease	: N.A.
(d) Name, address, and nationality of the publisher, if any, of the original work	: N.A.
(e) Name, address, and nationality of the publisher, or adaptation including the name, address and nationality of party authorizing	: N.A.

3. Remarks, if any

Place:
Date: **03/11/2023**

For : VEDANSHU THUNE

Proprietor