



Hand Gesture Recognition Using AI/ML

Tejasvi Jawalkar ¹, Sejal Sandeep Khalate ², Shweta Anil Medhe ³, Kshitija Shashikant Palaskar ⁴

Department of Computer Engineering
JSPM's Imperial College of Engineering and Research Pune, India

Abstract- Hand gesture recognition is a vital aspect of human-computer interaction, enabling natural and intuitive communication between users and machines. This paper presents a comprehensive review of recent advancements in the field of hand gesture recognition utilizing artificial intelligence (AI) and machine learning (ML) techniques. We discuss the challenges, methodologies, and applications associated with hand gesture recognition systems. Various AI and ML algorithms, including deep learning models such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), are explored for their efficacy in recognizing hand gestures from image or sensor data. Sign language is composed of continuous gestures, therefore, for sign language recognition, in addition to spatial domain, it also needs to capture motion information across multiple consecutive video frames. This paper explains two way communications between the deaf, dumb and normal people which means the proposed system is capable of converting the sign language to text and voice.

Keywords- Hand Gesture Recognition, Artificial Intelligence, Machine Learning, Convolutional Neural Networks, Recurrent Neural Networks, Human-Computer Interaction.

I. INTRODUCTION

In our daily lives, we encounter numerous individuals grappling with various forms of sensory impairments, hindering their ability to engage effectively with others. Historically, existing solutions predominantly hinged upon sensor-based mechanisms, yet these approaches have been characterized by their lack of universality and effectiveness. This article aims to introduce an innovative technology designed to revolutionize virtual communication for individuals with sensory impairments, offering a comprehensive solution without the need for sensors.

The newly developed technology heralds a paradigm shift in facilitating virtual conversations for individuals with sensory impairments. Unlike its predecessors, which heavily relied on sensor-driven methodologies, this groundbreaking approach transcends the limitations imposed by such mechanisms. By eschewing sensor dependency, this technology provides a more inclusive and versatile platform for communication, thereby enhancing accessibility and usability for individuals across diverse sensory profiles.

This innovative technology boasts several distinguishing features that set it apart from conventional sensor-based solutions. Firstly, it offers a sensor less framework, eliminating the need for cumbersome sensory apparatus. Secondly, it employs advanced algorithms and machine learning techniques to interpret and facilitate communication, thereby ensuring accuracy and reliability. Thirdly, it prioritizes user-friendliness and accessibility, catering to the diverse needs of individuals with sensory impairments. Lastly, it integrates seamlessly with existing communication platforms, enhancing interoperability and ease of adoption.

The adoption of this sensor less technology heralds' transformative benefits for individuals with sensory impairments and society at large. By democratizing virtual communication, it fosters inclusivity and equal participation, empowering individuals to engage meaningfully with their peers and communities. Moreover, it mitigates the barriers imposed by traditional sensor-based approaches, thereby enhancing the quality of life and societal integration for individuals with sensory impairments. Additionally, its scalability and versatility render it applicable across various domains, spanning education, healthcare, and social interaction.

With the proliferation of devices ranging from smartphones to virtual reality (VR) headsets, the need for seamless human-computer interaction has become increasingly paramount. Hand gesture recognition offers a promising solution by enabling users to convey commands or input without physical contact with the device. Traditional approaches to hand gesture recognition often relied on heuristic-based algorithms, which lacked robustness and adaptability to diverse environments and user variations. In contrast, AI and ML techniques empower systems to learn complex patterns and features directly from data, leading to more accurate and versatile recognition capabilities.

Sign language recognition methods are easily affected by human movement, change of gesture scale, small gesture area, complex background, illumination and so on. And some sign language recognition methods must use gesture areas to input information [1]. Therefore, robust hand locating is an important pretreatment step in sign language recognition. Compared with basic gestures, gestures in sign language are characterized by complex hand shape, blurred movement, low resolution of small target area, mutual occlusion of hands and faces, and overlapping of left and right hands. In addition to the influence of complex background and light, a large number of sign language image sequences are needed in sign language recognition, and all these have brought great challenges to the accuracy and stability of hand locating in sign language recognition. Sign language is composed of continuous gestures, therefore, for sign language recognition, in addition to spatial domain, it also needs to capture motion information across multiple consecutive video frames. At the same time, how to build an efficient and suitable sign language recognition model has always been a hot research spot.

I. PROPOSED SYSTEM

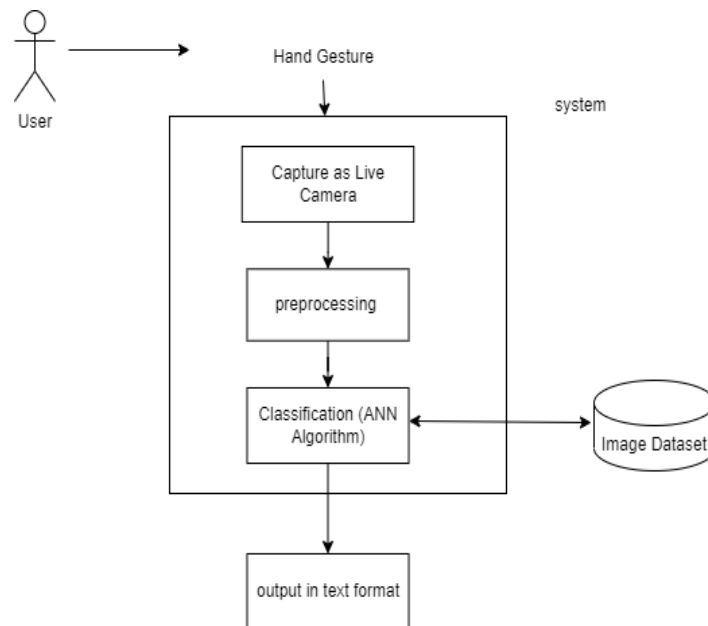


Figure 1: Proposed System of Hand Gesture Recognition

First, they will collect the dataset, then the machine has to process the dataset after training, then the machine has to train the data using ANN algorithm and then detect the output. So when you provide the input as a camera then detect the output [2].

Recognition of sign language is an emerging area of research in the domain of gesture recognition. Research has been carried out around the world on sign language recognition, for many sign languages. The basic stage of a sign language recognition system is accurate hand segmentation. This paper used Otsu's technique of segmentation to create an improved vision-based recognition of sign language. Approximately 466 million users worldwide suffer from hearing loss, of which 70 million suffer from hearing loss. The number of children is 34 million Recognition.

Deaf people have very little or no capacity to hear. For speech, they use the sign language. People in distinct areas of the globe use distinct sign languages, which are very small in number compared to spoken languages [3]. Our goal is to create a static-gesture recognizer, a multi-class classifier that predicates the gestures of the static sign language. In the proposed work, they identified the hand in the raw image and provided the static gesture recognizer (the multi-class classifier) with this section of the image. They first build the dataset and build a multi-class classifier from the scikit-learn library [3]. The easiest way to describe artificial neural networks is as computer models of biological systems used to do a certain set of tasks, such clustering, classification, pattern recognition, etc [2]. An artificial neural network is a network of artificial neurons with biological inspiration that is set up to carry out a predetermined set of tasks. Artificial neural networks, ANNs for short, have become well known and also considered as a hot topic of interest and are used in chatbots commonly used in text classification To be honest with you about, that you are only a neuroscientist, if you're a neuroscientist It wouldn't make much sense. Software for simulations of synapses and neurons in animal brains is evolving as the neural software industry has already been around for decades Nature is an inspiration to people. For example, bird-powered airplanes were developed. Similarly, artificial neural networks (ANNs) were developed in the brain using neurons. This approach has addressed challenging machine learning problems such as image classification, recommendation algorithms, and speech language interpretation. The ANN method is a machine learning algorithm with natural neural networks. ANN is the essence of deep learning. The first ANN was proposed in 1944, but has become increasingly popular in recent years. Let's take a look at why this method has become so popular in recent years. As you know, with the rise of the internet and social media, the amount of data generated has increased, and big data has become a buzzword. Big data has made it easier to train ANNs. While traditional machine learning algorithms rarely analyze big data, artificial neural networks performed well on big data. Another reason for the popularity of these algorithms is when gaming industry GPUs have increased the computing power of machines is that It was made. Using GPUs. Also, the development of awesome architectures like CNN, RNN, Transformers etc.

II. METHODOLOGY

AI and ML techniques have revolutionized hand gesture recognition by enabling the extraction of meaningful features from raw input data. Convolutional neural networks (CNNs) have demonstrated remarkable success in image-based gesture recognition tasks, leveraging hierarchical feature learning to capture spatial information effectively. Recurrent neural networks (RNNs) and their variants, such as long short-term memory (LSTM) networks, excel in sequential gesture recognition tasks, where temporal dynamics play a crucial role. Additionally, hybrid approaches combining CNNs and RNNs have emerged to leverage both spatial and temporal information for improved performance.

"Artificial neural networks" refers to a branch of artificial intelligence inspired by biology and based in the brain[1]. Computer networks based on the biological neural networks that make up the human brain system are often referred to as artificial neural networks Artificial neural networks also contain neurons that are connected at different levels of communication, much like neural networks in a real brain in the 19th century. These nodes are called nodes. Deaf people use hand signs to communicate, so normal people have trouble recognizing their speech with synthetic signs. Therefore, there is a need for systems that recognize signals and disseminate information to the general public. People with hearing loss always have trouble communicating with the general public. They worry about communicating their thoughts and ideas to the general public with very little or sometimes no knowledge of sign language.

- This results in community members with hearing loss losing interest in normal activities and sometimes avoiding contact with normal people and becoming isolated.
- Researchers have developed many sign language recognition systems to overcome this situation but there is still a need for accurate and effective signal recognition. Currently, the systems proposed by previous researchers are based on the transformation of a process-based process into an equivalent representation.
- These systems limit the maximum number of action verbs to be processed in a particular language.
- The aim of the study is to develop a sign language recognition system for English sounds.
- The proposed system should be user-friendly and user-friendly for people with hearing loss.

Multilayer perceptron's consist of an input layer, a hidden layer, and an output layer. As you can see in the image above, there's a hidden layer. If there is more than one hidden object, it is called a deep neural network. This is where deep learning comes into play. The development of modern AI architecture has made deep learning popular.

In short, inputs pass through neurons and predictions are made. However, how can prognosis of neural tubes be improved? This is where the backpropagation algorithm comes in. This algorithm assumes a neural.

The literature on artificial neural networks covers every aspect of these networks [2]. They will explore artificial neural networks (ANNs), adaptive resonance theory, and Kohonen's own systems.

IV.APPROACHES

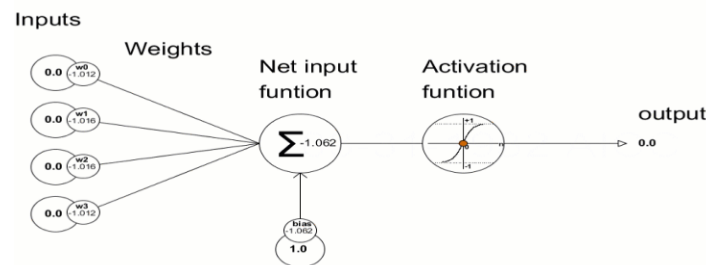


Figure 2: ANN ARCHITECTURE

How do artificial neural networks work?

An artificial neural network can be satisfactorily represented as a weighted directed graph, where the synthetic neurons shape the nodes [1]. The association between the neuron outputs and neuron inputs can be seen because the directed edges have weights. The artificial neural network receives the input signal from the outside source in the form of a pattern and a picture inside the form of a vector. These inputs are then mathematically assigned by way of the notation $x(n)$ for every n quantity of inputs [2].

Each input is then multiplied by its corresponding weight (this weight is the detail used by neural networks to solve a particular problem). In general, this weight usually represents the strength of connections between tissues in artificial tissues. All weighted input parameters are summarized in the software package [4].

If the weighted sum equals zero, a bias is added to make the output zero or otherwise scale up to the system's response. The bias has the same input and is weighted equal to 1. Here, all weighted inputs can range from 0 to positive infinity [5]. Here, a fixed maximum value is compared to keep the response within the desired value range, and all weighted inputs are passed through the processing function.

The activation function specifies the transfer functions used to achieve the desired output. There are special activation functions, but they are a linear or nonlinear set of functions. Some commonly used activation functions

are binary, linear, and tonal hyperbolic sigmoidal activation functions. input is then multiplied by its corresponding weights (these weights are the details used by artificial neurons to solve a particular problem), and in general, this weight usually represents the strength of the intermolecular connections in the artificial interface [1]. All weighted input parameters are summarized in the software package. If the weighted sum equals zero, a bias is added to make the output zero or something else to scale up to the system's response. The bias has the same input, and the weight is equal to 1. Here, all the weighted inputs can range from 0 to positive infinity [2]. Here, a fixed maximum value is compared to keep the response within the desired value range, and all weighted inputs are passed through the processing function.

The activation function specifies the transfer functions used to achieve the desired output. There are special activation functions, but they are a linear or nonlinear set of functions. Some commonly used activation functions are binary, linear, and tonal hyperbolic sigmoidal activation functions

V.OUTPUT

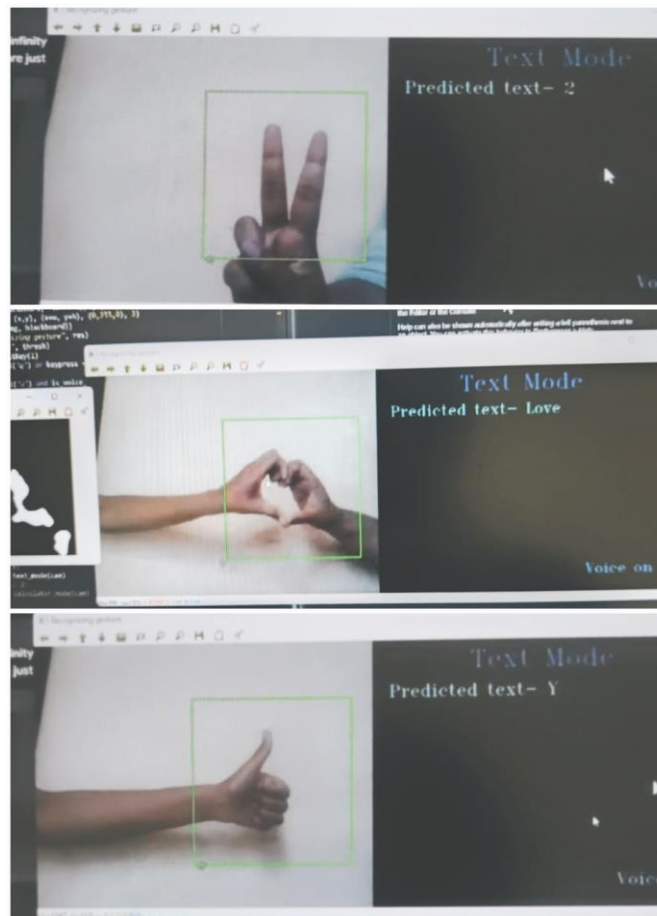


Figure 3

VI.CONCLUSION

In conclusion, the advent of this sensorless technology represents a watershed moment in the realm of virtual communication for individuals with sensory impairments. By transcending the limitations of sensor-based methodologies, it offers a comprehensive and inclusive solution, empowering individuals to communicate effectively and participate fully in society. Moving forward, continued research and development in this domain hold the promise of further advancing accessibility and inclusivity, ushering in a more equitable and empathetic society for all.

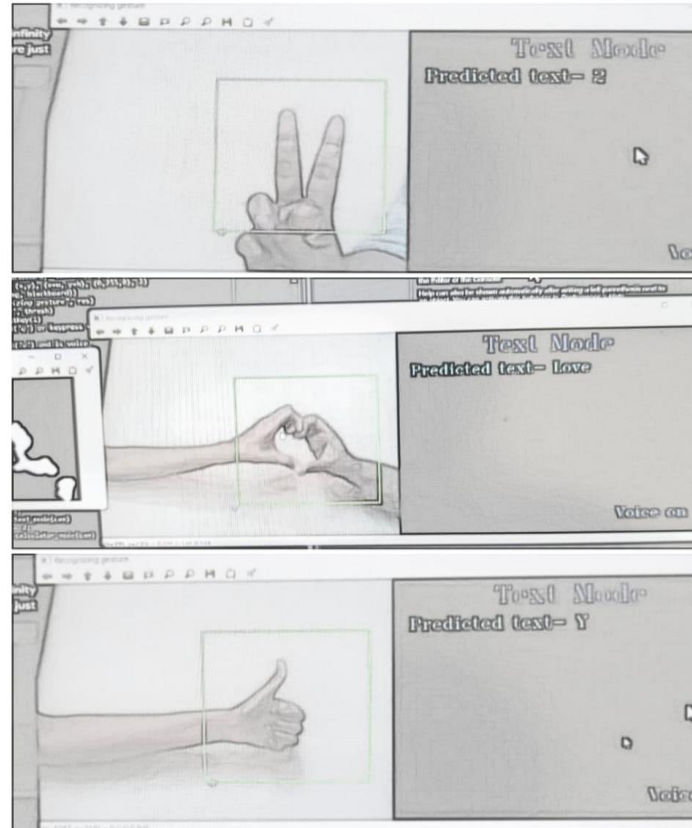


Figure 4

REFERENCES

- 1.C. Ong, W. Ibrahim, and S.M. Sapuan, "Recent Advances In Hand Gesture Recognition For Human-Computer Interaction: A Review," Computer Methods And Programs In Biomedicine, vol. 175, pp. 137-159, 2019.
2. Y. Lecun, f. Huang, and l. Bottou, "Learning Methods for Generic Object Recognition with Invariance to Pose and Lighting," Proceedings of The IEEE Conference On Computer Vision And Pattern Recognition (Cvpr), 2004.
3. C. Li, y. Tian, s. Zhang, and l. Wang, "Hand Gesture Recognition Based on Convolutional Neural Networks," Proceedings of The Ieee International Conference On Computer Vision (Iccv), 2017.
- 4.S. S. Malik and k. N. Jha, "hand gesture recognition techniques: a review," Journal of Ambient Intelligence and Humanized Computing, vol. 11, no. 1, pp. 311-340, 2020.