A Systematic Review on Manual Language Interpretation: Understanding the Sign Communication

Bhavana N, Govind Sreekar Shenoy
Department of Information Science,
Nitte Meenakshi Institute of Technology, Yelahanka, Bangalore, India

Abstract - This comprehensive literature review examines the implementation of Sign Language Recognition (SLR) for automated sign-language translation. While SLR has been extensively researched, its practical implementation remains rare due to complexity and resource requirements. The review analyzes various techniques and methodologies used in creating working models of sign-language translators, aiming to explore the potential of Artificial Intelligence and Machine Learning technology in enhancing automated American Sign Language translation. The analysis reveals notable achievements in selected research studies, yet acknowledges their inherent strengths and weaknesses. Additionally, the review covers deep learning techniques such as CNN, LSTM, and RNN, along with machine learning techniques like K Nearest Neighbor, PCA, LDA, and HMM, showcasing their relevance in classification, dimensionality reduction, and sequence modeling. The exploration encompasses both sensor-based and vision-based recognition methods, highlighting their applications and advantages. Emphasis is also placed on the pivotal role of feature extraction and the promising application of transformer models in sign language recognition. Overall, this review provides valuable insights into the advancements, strengths, limitations, and potential applications of deep learning methods for sign language recognition, essential for fostering effective communication between the deaf community and the hearing population.

Index Terms - SLR, CNN, LSTM, RNN, K Neareast Neigbor , HMM, SVM, PCA , Sensor based recognition, Vision Based Recognition

I. INTRODUCTION

Communication between human beings has long been reliant on language, which is passed down through generations and varies across societies. However, individuals with hearing impairments, known as deaf people, face challenges in communicating through spoken language. Deafness and hearing loss are prevalent worldwide, affecting over 1.5 billion people, with 430 million experiencing disabling hearing loss. By 2050, it is projected that the number of individuals with disabling hearing loss will surpass 700 million. Those who are hard of hearing, with varying degrees of hearing loss, often rely on Sign Language as their primary means of communication. Deaf, dumb, and hearing-impaired individuals encounter difficulties in interacting with others, but SLR bridges this gap, enabling them to communicate and engage with communication technology. Therefore, this review provides a exploration of Sign Language, gestures, and key techniques employed to differentiate Sign Language, while also delving into their translation using information technology and computer vision. These advancements facilitate interaction and communication with users, offering invaluable assistance in facilitating the seamless social integration of individuals with hearing impairments or deafness, enabling them to communicate naturally with others.

A. In Sign Language Recognition (SLR), the typical model steps involve:

1) Pre-processing: This includes image or video input acquisition, followed by noise reduction, image enhancement, and background removal to isolate the hand region
2) Feature Extraction: Relevant features are extracted from the hand region, which can include hand shape, hand position, motion trajectories, or hand landmarks using techniques like edge detection, skeletonization, or deep feature extraction.
3) Gesture Segmentation: The hand gestures are segmented into meaningful units or sequences, enabling the recognition of individual signs or sign phrases.
4) Classification: Machine learning or deep learning models are employed to classify the segmented gestures into their corresponding sign language categories or labels.
5) Post-processing: Further refinement and validation of the recognized gestures may be performed using techniques like temporal smoothing, post-classification rules, or language modeling to improve accuracy and coherence of the recognized sign language output.

II. LITERATURE SURVEY

A. REVIEW BASED ON DEEP LEARNING

Sign language recognition involves the interpretation and translation of hand gestures and movements into meaningful text or spoken language. Traditional approaches to sign language recognition relied on handcrafted features and ML algorithms. However, with emergence of the deep learning, it has shown remarkable performance in capturing complex spatiotemporal patterns inherent in sign language. This review presents an analysis of various deep learning techniques employed in sign language recognition. It discusses the effectiveness of CNNs in extracting discriminative visual features from sign language videos or images. Furthermore, it explores the utilization of RNNs, particularly LSTM to record temporal dependencies in sign language sequences. In this study, the integration of the skeletonization algorithm and convolutional neural network in the gesture detection algorithm greatly enhances detection accuracy

and robustness, even in challenging imaging angles and environmental conditions. This approach optimizes the skeleton algorithm layer-by-layer, addressing the influence of shooting angles, while employing deletion techniques to recognize gestures of the same type and extract crucial node information from hand frame diagrams. Leveraging the hand's spatial coordinate axis, the algorithm determines gesture direction, enabling effective segmentation and overcoming environmental influences. The proposed approach achieves an impressive recognition rate of 96.01% and outperforms SVM, dictionary learning sparse representation, and CNN methods. By employing a skeletonization algorithm to analyse spatial data and identify key hand nodes, limitations associated with limited viewing angles are overcome. Challenges in distinguishing similar gestures are addressed using a jump motion controller and bidirectional recurrent neural network (RNN) that precisely calculate angle changes during dynamic hand movements. The combined approach effectively mitigates imaging angle and environmental impact, resulting in improved accuracy and robustness, paving the way for more reliable gesture recognition systems. [1]

An American Sign Language (ASL) Translator is developed. The development process involves utilizing a pretrained Google Net architecture that has been trained. The model underwent training using the ILSVRC2012 dataset, alongside ASL datasets sourced from Surrey University and Massey University. The translation process begins by acquiring a user's signing video as input. Each frame in the video is then classified to determine the corresponding letter. The output of the system displays the most likely word based on the classification scores. To accomplish this, a Convolutional Neural Network (CNN) is employed for ASL letter classification. CNNs are particularly effective at classifying visual images, as they can learn features and their associated weights. In this case, a softmax based function is used for optimizing the objective function. Transfer Learning is utilized, where models are trained on a large dataset and then fine-tuned through initialization. The ASL Translator is designed to be accessible through a web application and laptop camera, using colour images. However, it is worth noting that the classification accuracy varies across different letters. Letters from 'a' to 'e' are classified more accurately than letters from 'a' to 'k'. This discrepancy in accuracy can be attributed to the limited dataset available for training and the variations in environmental conditions during video capture. [2]

This study utilizes a dataset consisting of 2000 American sign language images, which are divided into 1600 images for training and 400 images for testing, maintaining an 80:20 ratio. The data collection process involves recording hand movements using a webcam, followed by image processing techniques. To ensure accurate analysis, background detection and removal are performed using the HSV colour removal algorithm. Segmentation techniques are then applied to isolate the area of skin tone. Morphological operations, including dilation and erosion, are utilized with an elliptical kernel, and a mask is employed for further image processing. Additionally, the images are resized to ensure uniformity. From each frame, binary pixels are extracted, and a CNN utilized for the purposes of training and classification. The CNN architecture comprises three layers: three convolutional layers utilizing the Rectified Linear Unit (ReLU) activation function, three pooling layers, and a fully connected layer with the Softmax activation function. To assess the performance of the system, precision, recall, and Fmeasure are determined for each class. These metrics offer valuable insights into the system's accuracy when it comes to recognizing various gestures. Subsequently, based on the input gesture from the user, the system predicts the corresponding gesture and displays the results accordingly. By following this methodology, the system effectively trains on the dataset, performs gesture classification using CNNs, and provides accurate predictions for user input gestures. [3]

Text-to-speech translation is performed to obtain results from frame-separated sign language input video. Each frame is put into Open Pose, an open-source convolutional neural network used to estimate hand poses for each frame of videos. A long short-term memory network is then built under the Tensor flow framework to classify hand pose models. The spatial feature of every 30 continuous frames is stored as a matrix to extract the temporal feature of the LSTM network. And every other successive table gives an 80 percent overlap. Finally, the LSTM predicts a possible translation and presents it. Spatial and temporal features are separated. Regularization and truncation are used to reduce model over fitting. To train this network, a dataset containing 3060 videos of seventeen different words and phrases is created and augmented. With optimization, LSTM achieves an accuracy of 93.62 and sign language is recognized. [4]

This approach leverages a camera as the primary tool for manual tracking and employs deep learning techniques for gesture classification. The progress of the American Sign Language recognition system eliminates the need for specialized hardware beyond the camera. The system utilizes a fusion of CNN and LSTM models. The CNN model undergoes training to discern spatial features, while LSTM captures temporal features, enabling accurate detection of both static images and dynamic gestures. However, certain factors can affect the accuracy of the model. Variations in facial appearance and clothing worn by the user may introduce challenges. Additionally, differences in lighting conditions and skin tones can impact the model's performance. The process begins by capturing input through a webcam, and the input frames are then converted to grayscale or HSV colour space. These frames are appropriately scaled and transformed before being fed into a pre-trained CNN model. The CNN predicts the gesture, and the output is classified into the corresponding category, providing the desired recognition outcome. This approach allows for efficient ASL recognition using readily available equipment and deep learning techniques. [5]

A new dataset of 2340 samples from diverse backgrounds was curated specifically for Bangla alphabet recognition. Their customized CNN architecture surpassed state-of-the-art models (ResNet, Efficient Net InceptionV3, and VGG19) achieving an impressive accuracy of 92% on the Bangla alphabet dataset. This demonstrates the efficacy of their tailored CNN in addressing Bangla alphabet recognition challenges. Additionally, their research introduces a system capable of recognizing Indian Sign Language using the Visual Words model. It incorporates skin color segmentation, background subtraction, SURF features, and histograms for accurate gesture recognition. SVM and CNN classification techniques ensure efficient recognition, while a user-friendly GUI enhances usability and accessibility for effective communication. [6]

B. REVIEW BASED ON MACHINE LEARNING

Sign language recognition involves the interpretation and translation of sign language gestures into text or spoken language. Machine learning techniques have been widely applied to extract meaningful features from sign language data and classify them into corresponding gestures. This review aims to present an overview of the various machine learning methods employed in sign language recognition, highlighting their strengths and limitations. In this paper, various techniques are employed to detect different hand gestures in the system. These techniques include skin filtering, hand clipping, Canny Edge Detection, KL Transform, and classification based on angle and Euclidean distance. Canny Edge Detection is utilized due to its ability to address the false positive and false negative problems commonly encountered in edge detection. It achieves this by utilizing two thresholds: a lower threshold and a higher threshold. The lower threshold prevents false edges from appearing, while the higher threshold ensures that valid edge points are not lost. The Canny Edge Detection process consists of three steps to begin, the input image is subjected to smoothing through a Gaussian filter. Subsequently, the gradient magnitude and angular images are computed. Finally, non-maximum suppression technique is applied to the gradient magnitude image, and edge linking and detection are performed using connectivity analysis and dual thresholding. KL Transform is employed to eliminate correlated data, reduce dimensionality, minimize mean squared error, and provide good clustering properties. This transformation helps optimize the representation of data and enhances the efficiency of gesture classification. Skin filtering plays a crucial role in isolating the hand from the background. It involves capturing an image using a camera and converting the RGB image to the HSV color space. The resulting image is then filtered and smoothed, leading to the generation of a binary grayscale image. To ensure that only the desired hand image remains, a Binary Linked Object technique is employed to remove other objects with similar skin color in the background. After successful skin segmentation, hand clipping is performed to refine the hand region. Overall, these techniques work together to accurately detect and classify hand gestures, enabling effective communication with the system. [7]

This approach utilizes state-of-the-art techniques to achieve highly accurate fingerprint recognition across a wide range of vocabularies. It combines an advanced appearance descriptor with a vocabulary model based on Hidden Markov Models to achieve exceptional accuracy. The process of identifying individual letters involves manual segmentation, appearance description, and classification. To determine the category of pixels (manual or non-manual), the model incorporates three components in synergy: signer-specific skin color model, spatially variable non-skin color model, and spatial coherence prior. This approach effectively differentiates between pixels belonging to the signer's manual actions and those representing other elements. The manual segmentation model follows a bootstrap approach, starting with the detection of the signer's face. Automatic labels are assigned to image pixels, distinguishing the face, clothes, and background using color cues such as red, green, and blue. A spatial prior is then learned, and a mask is trained to eliminate the facial region. For hand images, the hand shape is depicted by transforming histogram of oriented gradients (HOG) into a composite histogram. In word recognition, the classifier is fused or combined with a dictionary of known words, resulting in the creation of a hidden Markov model. This model helps refine the classifier's output and resolves ambiguities presented by visually challenging characters. Two primary descriptors are employed to detect individual letters. The GH descriptor, originally proposed by Goh and Holden, serves as a portrait descriptor. On the other hand, the OH descriptor is specifically designed for hand gesture detection and relies on the orientation histogram. By leveraging these advanced techniques and models, this approach achieves exceptional accuracy in fingerprint recognition and letter classification across extensive vocabularies. [8]

The proposed methodology comprises several crucial steps, including segmentation, feature extraction, visual vocabulary histogram generation, and classification. The process begins with image feature extraction as a preprocessing technique. Subsequently, a fast and efficient feature detection method called ORB (Oriented FAST and Rotated BRIEF) is applied. The ORB approach involves segmenting the image using an edge detection method and extracting features using ORB. To construct a feature bag model encompassing all descriptors, the K-means clustering algorithm is utilized. This algorithm clusters the extracted features into groups, forming the visual vocabulary. The descriptors are assigned to different clusters, and a histogram is generated based on the frequency of occurrence of each cluster in the image. For classification, various classifiers are employed, including Random Forest, SVM, Naive Bayes, Logistic Regression, Multilayer Perceptron, and KNN. The images are trained using these classifiers to learn the patterns and relationships between the extracted features and their corresponding classes. The accuracy of each model is then calculated using the training set, allowing for performance evaluation. By following this methodology, the system can effectively segment images, extract relevant features, generate visual vocabulary histograms, and classify images using different classifiers. This comprehensive approach enhances the accuracy and performance of the system in various visual recognition tasks. [9]

The proposed method represents a significant advancement in the development of a highly accurate system while reducing complexity. This paper introduces a hand gesture recognition system specifically designed to identify dynamic gestures performed against complex backgrounds. Notably, this novel approach eliminates the need for an instrumented glove or any additional markers, relying solely on 2D video input captured from bare hands. The resulting motion data is then leveraged for gesture recognition. To achieve real-time image recognition, this project employs principal component analysis and linear discriminant analysis algorithms for feature extraction from a training database. These extracted features are subsequently classified using the K Nearest Neighbor algorithm. The system utilizes a web camera for capturing live images, while the training set undergoes preprocessing. Real-time image preprocessing involves applying PCA and LDA algorithms, and the final classification is performed using the KNN algorithm. By utilizing this methodology, the system achieves real-time hand gesture recognition without the need for additional equipment or markers. The combination of PCA, LDA, and KNN enables efficient feature extraction and accurate classification. [10]

This approach enables the development of precise systems for recognizing hand gestures in complex dynamic environments. The article focuses on an automatic translation system that recognizes static gestures in the alphabet and American Sign Language (ASL). By combining the Hough transform and trained neural networks, gloves or visual cues are not required, allowing for natural interaction using bare hands. The image undergoes processing, converting it into a feature vector representation for recognition and classification. Techniques such as Canny Edge detection and the Hough Transform extract relevant information, facilitating accurate gesture recognition. The system achieves an impressive accuracy rate of 92.3% in recognizing ASL signs, demonstrating its potential in facilitating communication for sign language users. [11]

This research paper introduces a method for recognizing finger spelling in the American Sign Language alphabet utilizing the k-Nearest Neighbors classifier. The study further investigates the effect of dimensionality reduction using Principal Component Analysis (PCA) on the performance of k-NN classifier. The empirical findings demonstrate that k-NN classifier attains its peak accuracy of 99.8 percent with k = 3 when utilizing the complete set of features. However, when the model is represented by reduced dimensions obtained through PCA, the accuracy of the k-NN classifier significantly drops to 28.6 percent for k = 5. This decline in accuracy can be attributed to various factors, including the presence of redundant or highly correlated features within American Sign Language alphabet dataset. These characteristics pose challenges for PCA in effectively separating and distinguishing the data. Despite the decreased accuracy observed with PCA dimensionality reduction, the KNN classifier remains well-suited for applications in early childhood education, especially for the development of self-assessment systems catering to students with special needs who are learning ASL alphabet fingerspelling. In summary, this research paper provides valuable insights into finger spelling recognition in ASL using the k-NN classifier. It highlights the influence of PCA on dimensionality reduction and underscores the potential of this approach in educational contexts, particularly for ASL alphabet learning among students with special needs. [12]

This study focuses on ASL recognition using Kinect sign data and explores the effectiveness of the dynamic time warping (DTW) approach. They conducted an experiment to measure the similarity of sign trajectories using DTW and represented hand shapes using the Oriented Gradient Histogram (HoG) technique. The results of their experiment shows an improvement over previous work, achieving an accuracy of 82% with 10 matches, highlighting the effectiveness of DTW and HoG for ASL recognition. In addition to refining character detection accuracy, they put forward a simple RGB-D aligner that approximates alignment parameters between color (RGB) and depth frames. This aligner facilitates the synchronization of information from different modalities, leading to enhanced overall system performance. By combining DTW, HoG, and their proposed RGB-D aligner, they contributed to the advancement of ASL recognition technology. Their approach demonstrates promising results in accurately interpreting and recognizing sign language gestures, which in turn promotes more accessible and efficient communication for individuals using ASL. The combination of DTW, HoG, and the RGB-D aligner provides a robust framework for ASL recognition, creating opportunities for enhanced communication and comprehension between sign language users and individuals who do not use sign language. [13]

This paper introduces a technique for recognizing alphabets and numbers in American Sign Language using image saliency detection. This proposed approach involves feature detection, followed by processing the images using PCA and LDA to reduce dimensionality as well as enhances intra-class similarity while minimizing extra-class similarity. The extracted feature vectors undergo training and classification using neural networks. This system aims to enhance communication with the deaf community and establish a connection with computers through the use of standard sign language letters. Performance evaluation experiments were conducted using a standardized dataset, resulting in an average recognition accuracy of 99.88% across four training cycles using quadruple cross-validation. The results highlight the system's exceptional accuracy and outperformance compared to alternative methods, showcasing its potential in effectively recognizing ASL alphabets and numbers, thereby improving communication and accessibility for sign language users. [14]

C. Review based on Sensor Based Recognition

This study proposes a multimodal Kinect system for language learning in deaf children and compares it to the CopyCat system. The CopyCat system, utilizing colored gloves and accelerometers, faces challenges with maintenance, battery charging, and glove washing. To overcome these issues, the proposed system utilizes the Microsoft Kinect sensor, enhancing user comfort, durability, and ease of use. Depth information and RGB image streams are used for real-time signing and validation of American Sign Language (ASL) sentences. The system incorporates RADAR, LIDAR, structured light technology, and a compositor for depth map generation. Data collection involves comparing the performance of CopyCat and Kinect frameworks through sentence repetition tasks and sitting/standing kinetic data. Feature extraction, hidden Markov model training, and generalization measurement are key steps in the experimental design. [15]

This paper presents two distinct systems for recognizing ASL words, both utilizing ANN to translate them to English. The first system utilizes feature vectors captured at five different time points, while the other system/model employs histograms of feature vectors. The extraction in gesture features involves the use of sensory devices, specifically the motion tracking Flock of Birds 3-D and Cyberglove. The extensometers on Cyberglove provide data on knuckle angles, determining the hand shape, while the motion tracking device captures the hand's trajectory. These two systems process the data obtained from devices using both NN: speed and word recognition network. The speed network analyses the input data to determine duration of manual words. The sign gestures are demonstrated by feature vectors that encompass various aspects, including hand shape, position, direction, movement, bounding box, and distance. The other network functions as classifier, converting ASL signs to words depending on their features or histograms , the performance of systems are evaluated such that, they were trained and tested on a dataset comprising 60 ASL words, with varying numbers of samples. Comparative analysis was conducted between the two methods. The test results demonstrate detection accuracies of 92% and 95% for the two respective systems. By utilizing artificial neural networks and incorporating sensory devices for feature extraction, these systems provide an effective approach to recognize ASL words and translate them into English. The combination of feature vectors and histograms enables accurate classification and improves recognition performance. The evaluation results showcase the systems' high detection accuracies, emphasizing their potential in practical ASL word recognition application. [16]

Hand gesture recognition has gained significant popularity, especially with the use of sensors such as the (LMC) or Kinect sensor. In one particular system proposed, the combination of SVM and LMC was utilized for recognizing the American Sign Language (ASL) alphabet. The LMC sensor served as the input for capturing various hand signals, and a Deep Neural Network (DNN) model was employed, demonstrating superior accuracy compared to SVM. The study highlighted the importance of the relative distance between the tips of adjacent fingers in character recognition. However, a drawback of this approach was the relatively lower prediction accuracy due to the high similarity between certain letters and numbers. In another study, a Microsoft Kinect sensor was utilized for hand gesture recognition, taking advantage of its ability to detect depth and location information. This sensor provided valuable data that could be used in the recognition process. The utilization of sensors like LMC and Kinect in hand gesture recognition systems has opened up new

possibilities for accurate recognition and interpretation of gestures. The integration of machine learning models, such as DNN and SVM, with these sensors allows for efficient recognition and classification of hand gestures. Despite certain limitations, ongoing research in this field continues to advance the accuracy and applicability of hand gesture recognition systems. [17]

This study introduces a translator based on dynamic tense formation, which enhances the coordination and matching of prefixed words in the database. This enables the display of text pronunciation and its corresponding input character. To capture sign language gestures, they utilized Microsoft's Kinect sensor. Their comprehensive dataset consists of 30 distinct dictionaries, encompassing self-made signs in Standard Arabic Sign Language. The system operates in various modes, including online mode, signer-dependent mode, and signer-independent mode, allowing signers to express tags freely and naturally. Through extensive experimentation using real data collected from Arabic sign language, they demonstrated that their system outperforms others in terms of recognition accuracy across all modes. In signer-dependent network cases, the system achieves an impressive detection rate of 97.58 percent. Furthermore, in signer-independent network cases, the system achieves a commendable detection rate of 95.25 percent. The translator proposed in this study combines dynamic tense formation and Microsoft's Kinect sensor to provide an efficient and accurate solution for sign language translation. Its superior recognition performance, along with its flexibility in accommodating different signers, makes it a valuable tool for facilitating effective communication in Arabic sign language. [18]

The emergence of affordable depth sensors, such as the Leap Motion controller and Microsoft Kinect sensor, has opened up new avenues for human-computer interaction (HCI). This article introduces a novel multisensory federated framework for sign language recognition (SLR) that utilizes a coupled hidden Markov model (CHMM). Unlike traditional HMMs that rely on observation states, the CHMM operates in spatial space, allowing for the modeling of intermodal dependencies and enhancing interaction. The framework presented in this study specifically focuses on detecting dynamic isolated gestures in individuals with hearing impairment. To assess its effectiveness, the dataset was evaluated using various existing data fusion techniques. Significantly, the CHMM based approach achieved the highest accuracy in detection, reaching an impressive 90.80%. This substantial improvement in performance surpasses the outcomes obtained by popular data fusion methods commonly employed in the field. This research highlights the potential of the multisensory federated framework with CHMM for sign language recognition. By incorporating spatial information and capturing intermodal dependencies, the proposed approach offers a more accurate and robust solution for gesture recognition in SLR applications. [19]

The sign language recognition system (SLR) revolutionizes communication for individuals with hearing impairments. This study employs surface electromyography (sEMG) technology to precisely recognize the American Sign Language (ASL) alphabet. Capturing 27 ASL gestures, sEMG data is gathered from the right forearm. Feature extraction involves time and frequency domain analysis, specifically focusing on band power. Classification employs support vector machines and ensemble learning algorithms, ensuring accurate ASL interpretation. This integration of sEMG data and advanced classification techniques empowers users to construct words and sentences with ease. The SLR system has the potential to bridge communication gaps and foster inclusivity for individuals with hearing impairments, enhancing their quality of life. [20]

This research presents a system for recognizing American Sign Language (ASL) utilizing a compact and cost-effective 3D motion sensor, the palm-sized Leap Motion sensor. The use of this sensor offers a portable and affordable alternative to previous studies' reliance on the Cyberglove or Microsoft Kinect devices. The system employs the k-nearest neighbor and support vector machine algorithms to classify the 26 letters of the English alphabet in ASL, leveraging sensory data-derived features. Experimental evaluation reveals that the k-nearest neighbor algorithm achieves an average classification rate of 72.78%, while the support vector machine algorithm achieves 79.83%. The paper also includes in-depth discussions on parameterization of machine learning methods and the accuracy of specific alphabet letters. These discussions shed light on considerations for optimizing machine learning approaches and addressing the accuracy of individual alphabet letters Overall, this research showcases the feasibility of the proposed ASL recognition system, which leverages a compact and affordable 3D motion sensor. By utilizing machine learning algorithms and analyzing sensory data, the system demonstrates promising results in recognizing ASL alphabet letters. The findings from this study provide insights into optimizing machine learning parameters and improving accuracy for specific ASL gestures. [21]

This research paper presents a novel algorithm that utilizes the Kinect sensor for modeling and recognizing sign language. The main hypothesis is that certain frames within sign language videos possess both discriminative and representative qualities. Building upon this hypothesis, the algorithm assigns a binary latent variable to each frame in the training video, indicating its level of relevance. Subsequently, a latent support vector machine model is developed to classify characters and identify frames that are highly characteristic and representative in the videos. In addition, to further improve detection accuracy, the algorithm combines the depth map and color image captured by the Kinect sensor. By integrating these two modalities, a more efficient and precise function is created, resulting in enhanced recognition performance. To evaluate the proposed approach, a comprehensive dataset of American Sign Language consisting of approximately 2,000 sentences was collected. Each sentence was recorded using the Kinect sensor, capturing color, depth, and skeleton information. Through experiments conducted on this dataset, the effectiveness of the proposed sign language recognition method is successfully demonstrated. Overall, this research introduces an innovative algorithm that leverages the Kinect sensor to model and recognize sign language. By considering the relevance of frames and integrating depth and color information, the proposed approach achieves notable advancements in sign language recognition. [22]

The aim of this article is to enable the translation of American Sign Language (ASL) from static postures. To accomplish this objective, the researchers have developed a specialized glove embedded with six distinct colored markers and have devised an algorithm for alphabetic classification. The system is equipped with two cameras to capture the 3D coordinate points of each marker, enabling precise detection and analysis. The algorithm implemented in this research consists of three primary processes. Firstly, the Circle Hough Transform is employed for feature detection, allowing the identification of markers based on their distinctive circular shape. Next, the algorithm calculates all possible states of the triangular area using the 3D coordinate triple, which serves as a novel feature for characterization. Finally, a neural network is utilized for feature classification, providing valuable feedback for accurate recognition.

The experimental results highlight the effectiveness of the proposed method, with an average accuracy rate of 95%. This demonstrates the high performance and feasibility of the system in accurately translating ASL from static postures. The combination of the specialized glove, marker detection, and neural network classification showcases the potential for effective communication between sign language users and non-sign language speakers. This research paves the way for improved accessibility and inclusivity by bridging the communication gap between individuals who use sign language and those who do not. [23]

D. Review Based on Vision Based recognition

This article introduces the Transformer Encoder as a highly effective sign language recognizer, specifically focusing on accurately recognizing static Indian characters. The authors propose a vision converter method to achieve this goal, which demonstrates significant performance improvements compared to other state-of-the-art convolutional architectures. In the context of static Indian sign language recognition, the proposed method adopts a unique approach. The input signal is divided into a sequence of positional markers, which are then passed through a transformer block consisting of four self-attention layers and a multi-layer perceptron network. This architecture enables the model to capture essential features and dependencies within the input sequence. Through experiments and evaluations, the results highlight the successful recognition of sign language gestures using the proposed method. The performance is further enhanced through the utilization of various augmentation techniques, ensuring robustness and generalization of the model. Remarkably, the proposed approach achieves an impressive accuracy rate of 99.29 percent, even with minimal training sessions, emphasizing its efficiency and effectiveness in static Indian sign language recognition. Overall, this article showcases the effectiveness of the Transformer Encoder and the proposed vision converter method for accurate recognition of static Indian sign language characters. The remarkable accuracy achieved demonstrates its potential as a powerful tool in sign language recognition applications. [24]

To develop a highly accurate vision-based gesture recognition system, it is essential to find an effective method for representing hand movements. This article presents a novel approach that utilizes the Hilbert spacefilling curve to represent hand images. The proposed method involves the segmentation of the hand, followed by the extraction of a feature vector using the Hilbert curve. Gesture classification is then performed using classifiers such as Support Vector Machine (SVM) and Random Forest (RF). The choice of utilizing the Hilbert curve representation is motivated by its effectiveness in capturing shapes on smooth backgrounds while preserving pixel localization. Moreover, this representation demonstrates invariance to translation, scale, and stretching, making it highly suitable for hand gesture recognition. These properties contribute to the robustness and accuracy of the proposed approach. Experimental results confirm the effectiveness and practicality of the proposed method, demonstrating its superiority over other real-time methods in the field of hand gesture recognition. The use of the Hilbert space-filling curve as a hand image representation technique offers significant advantages in accurately capturing and classifying hand gestures. [25]

III. PROPOSED WORK

The proposed work aims to explore and develop new deep learning techniques for Sign Language Recognition (SLR) to improve the accuracy and efficiency of existing systems. The following deep learning techniques will be investigated, Convolutional Neural Networks (CNNs): CNNs have shown great success in image classification tasks and can be adapted to SLR. Different CNN architectures, such as VGG, ResNet, or EfficientNet, will be explored and optimized for SLR to capture spatial features from sign language gesture images. Recurrent Neural Networks (RNNs): RNNs, particularly Long Short-Term Memory (LSTM) or Gated Recurrent Unit (GRU) models, are suitable for capturing the temporal dependencies in sign language sequences. They will be used to process sequential data, such as the temporal evolution of hand movements in SLR. Attention Mechanisms: Attention mechanisms enable models to focus on relevant parts of the input sequence while processing the data. By incorporating attention mechanisms into the network architecture, the SLR system can better understand and interpret crucial aspects of sign language gestures. Transformer Models: Transformer models have achieved significant breakthroughs in natural language processing tasks. Adapting transformer architectures, such as the popular BERT or GPT models, to SLR can capture the semantic meaning of sign language gestures and improve translation accuracy. Data Augmentation: Augmenting the available sign language dataset with various transformations, such as rotation, scaling, or cropping, can increase the diversity and robustness of the training data. This technique helps to reduce overfitting and improve the generalization ability of the SLR models. Transfer Learning: Pretraining deep learning models on largescale datasets, such as ImageNet or COCO, and finetuning them on sign language data can leverage the learned visual representations for SLR tasks. Transfer learning can enhance the performance of SLR models, especially when the available sign language dataset is limited. Model Ensemble: Ensemble methods, such as model averaging or stacking, can combine the predictions of multiple SLR models to obtain a more accurate and robust overall prediction. Ensemble techniques will be explored. This work will involve extensive experimentation and evaluation on benchmark SLR datasets, comparing the performance of the proposed techniques against existing approaches. The aim is to advance the field of SLR and contribute to the development of more accurate and efficient sign language recognition systems.

IV. Conclusions

In conclusion, sign language recognition is a ground-breaking technology that has the potential to revolutionize communication for individuals who are deaf or hard of hearing. By leveraging advanced computer vision, deep learning and machine learning techniques, researchers and engineers have made significant strides in developing systems that can accurately interpret and translate sign language gestures into text or spoken language. The advent of sign language recognition holds immense promise for improvising the gap between the deaf and hearing communities. It enables deaf individuals to express themselves more effectively in a variety of contexts, including educational, professional, and social settings. Moreover, it promotes inclusivity and equal access to information and services by facilitating real-time communication between deaf and hearing individuals. Despite the progress made, sign language recognition systems still face challenges, such as dealing with the complexity and variability of sign language across different regions and cultures. Further research and development efforts are needed to improve the accuracy, robustness, and versatility of these systems. In conclusion, sign language recognition represents a significant advancement in assistive technology, empowering individuals with hearing

impairments to engage in seamless communication with the world around them. As the technology continues to evolve, it holds the potential to transform lives, enhance accessibility, and foster greater inclusivity in society.

V. REFERENCES

[1] "Gesture recognition based on skeletonization algorithm and CNN with ASL database" Du Jiang & Gongfa Li & Ying Sun & Jianyi Kong & Bo Tao

[2] "Real-time American sign language recognition with convolutional neural networks" Brandon Garcia and Sigberto Alarcon Viesca

[3] "Sign Language Recognition System Using Convolutional Neural Network And Computer Vision " Romala Sri Lakshmi Murali L.D.Ramayya, V. Anil Santosh

[4] "A Sign Language Translation System Based on Deep Learning" Siming he

[5] "Sign Language recognition using deep learning and computer vision" R.S. Sabeenian S. Sai Bharathwaj M. Mohamed Aadhi

[6] "Sign Language Recognition for bangla alphabets using deep learning methods " Md. Saiful Islam Dhrubajyoti Das Saurav Das; Md. Nahid Ullah

[7] "Hand Gesture Recognition based on karhunen-Loeve Transform" Joyeeta Singha, Karen Das

[8] "Automatic recognition of fingerspelled words in British Sign Language" Stephan Liwicki & Mark Everingham

[9] "Hand Gesture Recognition using Image Processing and Feature Extraction Techniques " Ashish Sharma, Anmol Mittal, Savitoj Singh,Vasudev Awatramani

[10] "Vision Based Hand Gesture Recognition for Indian Sign Language" Ravikiran P Reehan Ahmed Jagadeesh B

[11] "American sign language (ASL) recognition based on Hough transform and neural networks "M. Qutaishat, Moussa Habeeb,

[12] Aryanie D, Heryadi Y (2015) "American Sign Language-based finger- spelling recognition using knearest neighbors classifier."

[13] "Sign Language Recognition using dynamic time warping and hand shape distance based on histogram of oriented gradient features" Pat Jangyodsuk Christopher Conly Vassilis Athitsos

[14] "Saliency based alphabet and numbers of American sign language recognition using linear feature extraction" Majid Zamani Hamidreza Rashidy Kanan

[15] "American Sign Language Recognition with the Kinect "Zahoor Zafrulla, Helene Brashear , Thad Starner, Harley Hamilton, Peter Presti

[16] "Linguistic properties based on American Sign Language isolated word recognition with artificial neural networks using a sensory glove and motion tracker"Cemil Oz , Ming C. Leu

[17] "American Sign Language Recognition Using Leap Motion Controller with Machine Learning Approach "Teak-Wei Chong and Boon-Giin Lee 2018

[18] Sarhan NA, El-Sonbaty Y, Youssef SM (2015) "HMM-based Arabic Sign Language recognition using Kinect "

[19] "Coupled HMM-based multi-sensor data fusion for sign language recognition" Pradeep Kumar, Himaanshu Gauba

[20] Wu J,Tian Z, Sun L, Estevez L, Jafari R (2015) "Real Time American Sign Language Recognition using wrist worn motion and surface EMG sensors"

[21] Chuan CH, Regina E, Guardino C (2014) American Sign Language recognition using leap motion sensor.

[22] "Latent support vector machine for sign language recognition with Kinect" Chao Sun; Tianzhu Zhang; Bing-Kun Bao; Changsheng Xu

[23] "American Sign Language Recognition by using 3D Geometric Invariant Feature and ANN Classification" Watcharin Tangsuksant, Suchin Adhan, Chuchart Pintavirooj

[24] "DeepVision Transformer for Sign Language Recognition" deep r. kothadiya , chintan m. bhatt , tanzila saba , amjad rehman and saeed ali bahaj

[25] "Sign Language Recognition Using Hilbert Curve Feature" Amira Ragab, Maher Ahmed, Siu Cheung Chau

[26] "Indian Sign Language recognition system using SURF with SVM and CNN" Sarhan NA, El-Sonbaty Y, Youssef SM (2015) HMMbased Arabic Sign Language recognition using Kinect.

[27] Deafness and hearing loss.2020.Available from: https://www.who.int/news-room/fact-sheets/detail/deafnessandhearing- loss.

[28] Tietze K. Clinical Skills for Pharmacists 3rd Edition; 2011. 3. [29] Ohna SE. Open your eyes: deaf studies talking.

Scandinavian Journal of Disability Research. 2010 Jun; 12(2): p. 141-146.

[30] Dumay J, Bernardi C, Guthrie J, Demartini P. Integrated reporting: A structured literature review. Accounting Forum. 2016 Sep 22; 40(3): p. 166-185.

[31] Ávila-Pesántez D, Rivera LA, Alban MS. Approaches for Serious Game Design: A Systematic Literature Review;2017.Available:http://www.asee.org/documents/papersandpublications/papers/CoEd_Journal2017/JulSep/AVILA_PESÁNTEZ.pdf.

[32] Joshi A, Sierra H, Arzuaga E. American sign language translation using edge detection and cross correlation. 2017 IEEE Colombian Conference on Communications and Computing, COLCOM 2017 - Proceedings. 2017.

[33] Jin CM, Omar Z, Jaward MH. A mobile application of American sign language translation via image processing algorithms. Proceedings - 2016 IEEE Region 10 Symposium, TENSYMP 2016. 2016;: p. 104-109.

[34] Guo D, Zhou W, Wang M, Li H. Sign language recognition based on adaptive HMMS with data augmentation. Proceedings - International Conference on Image Processing, ICIP. 2016; 2016Augus: p. 28762880.

[35] Aly W, Aly S, Almotairi S. User-independent american sign language alphabet recognition based on depth image and PCANet features. IEEE Access. 2019; 7: p. 123138-123150.

[36] Pigou L, Dieleman S, Kindermans PJ, Schrauwen B. Sign Language Recognition Using Convolutional Neural Networks. In Pigou L, Dieleman S, Kindermans PJ, Schrauwen B..; 2015. p. 572-578.

[37] Abdelnasser H, Harras KA, Youssef M. WiGest demo: A ubiquitous WiFi-based gesture recognition system. Proceedings - IEEE INFOCOM. 2015; 2015-Augus: p. 17-18.

[38] Jalal MA, Chen R, Moore RK, Mihaylova L. American Sign Language Posture Understanding with Deep Neural Networks. 2018 21st International Conference on Information Fusion, FUSION 2018. 2018;: p. 573-579.

[39] Brandon Garcia , Sigberto Alarcon Viesca. Real-time American Sign Language Recognition with ConvolutionalNeuralNetworks. [40] Rao GA, Kishore PVV. Sign language recognition system simulated for video captured with smart phone front camera. International Journal of Electrical and Computer Engineering. 2016; 6(5): p. 2176-2187

[41] Cayamcela MEM, Lim W. Fine-tuning a pre-trained Convolutional Neural Network Model to translate American Sign Language in Real time. 2019 International Conference on Computing, Networking and Communications, ICNC 2019. 2019;: p. 100-104.

[42] Shahriar S, Siddiquee A, Islam T, Ghosh A, Chakraborty R, Khan AI, et al. Real-Time American Sign Language Recognition Using Skin Segmentation and Image Category Classification with Convolutional Neural Network and Deep Learning. IEEE Region 10 Annual International Conference, Proceedings/TENCON. 2019; 2018(October): p. 1168-1171. 17. Thongtawee A, Pinsanoh O, Kitjaidure Y. A Novel Feature Extraction for American Sign Language Recognition Using Webcam. BMEiCON 2018 - 11th Biomedical Engineering International Conference. 2019;: p. 1-5.

[43] Yeo HS, Lee BG, Lim H. Hand tracking and gesture recognition system for human-computer interaction using low-cost hardware. Multimedia Tools and Applications. 2015; 74(8): p. 2687-2715.

[44] Huang J, Zhou W, Zhang Q, Li H, Li W. Video-based sign language recognition without temporal segmentation. 32nd AAAI Conference on Artificial Intelligence, AAAI 2018. 2018;: p. 2257-2264.

[45] Kumar P, Gauba H, Roy PP, Dogra DP. Coupled HMM-based multisensor data fusion for sign language recognition. Pattern Recognition Letters. 2017; 86: p. 1-8.

[46] Flores CJL, Cutipa AEG, Enciso RL. Application of convolutional neural networks for static hand gestures recognition under different invariant features. Proceedings of the 2017 IEEE 24th International Congress on Electronics, Electrical Engineering and Computing, INTERCON 2017. 2017;: p. 5-8.

[47] Taskiran M, Killioglu M, Kahraman N. A Real-Time System for Recognition of American Sign Language by using Deep Learning. 2018 41st International Conference on Telecommunications and Signal Processing, TSP 2018. 2018;: p. 1-5. 23. Xu P. A Real-time Hand Gesture Recognition and Human- Computer Interaction System. 2017 Apr 24.

[48] Ahmed W, Chanda K, Mitra S. Vision based Hand Gesture Recognition using Dynamic Time Warping for Indian Sign Language. In 2016 International Conference on Information Science (ICIS); 2016: IEEE. p. 120-125

[49] Cui R, Liu H, Zhang C. Recurrent convolutional neural networks for continuous sign language recognition by staged optimization. Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017. 2017; 2017-Janua: p. 16101618.

[50] Koller O, Zargaran S, Ney H, Bowden R. Deep sign: Hybrid CNNHMM for continuous sign language recognition. British Machine Vision Conference 2016, BMVC 2016. 2016; 2016Septe: p. 136.1-136.12. 27. Hore S, Chatterjee S, Santhi V, Dey N, Ashour AS, Balas VE, et al. Optimized Neural Networks. 2017;: p. 139-15